MATH 365 Stochastic Processes

Goals: (1) gain some intuition about stochasticity & randomness
(2) learn to simulate & analyze random processes

We'll begin with finite-state, discrete time processes, modeled as a sequence of random variables $X_0, X_1, X_2, \cdots$

Example: Bernoulli process X_n = result of nth coin flip
(1 for heads, 0 for tails)

(R demo)

State space Ω = all binary sequences

Example: Simple random walk $S_n = X_1 + X_2 + \cdots + X_n$, $X_k = \begin{cases} 1 & \text{heads} \\ -1 & \text{tails} \end{cases}$

(R demo)

Heads $\rightarrow$ move up
Tails $\rightarrow$ move down

<u>Markov chain</u>: X_{n+1} depends on X_n but no other previous X_k

Knowing $\mathbb{P}\{X_{n+1} = j \mid X_n = i\}$ for all states i & j, $n \geq 0$, characterizes the MC

Brief probability review

- State space Ω = set of all outcomes of a random experiment (possible states the stochastic process can take)

Event A = subset of Ω

Conditional probability of event A occurring given that event B occurred

$$\mathbb{P}(A \mid B) = P(A \cap B) / P(B)$$

A & B are <u>independent</u> if $P(A \mid B) = P(A)$

Law of Total Probability: If $B_1, \cdots, B_k$ are disjoint events and their union is all of Ω, then for any event A

$$\mathbb{P}(A) = \sum_{i=1}^{k} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{k} P(A \mid B_i) P(B_i)$$

→ Example: 7% of men are colorblind and 0.4% of women are colorblind.
49% of population is male, 51% female.
What is prob a randomly selected person is colorblind?

$$\mathbb{P}(\text{colorblind}) = P(C \mid M)P(M) + P(C \mid F)P(F) = (.07)(.49) + (.004)(.51)$$
$$= .03634$$

Bayes Rule $\mathbb{P}(B|A) = \dfrac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)}$ $\leftarrow \mathbb{P}(A)$

complement
(B didn't occur)

→ Example: Suppose one person in a 100-person company is a thief.
Company administers a polygraph test to its employees.
If subject is a liar, there is 95% chance polygraph will detect this.
If subject is truthful, there is 10% chance of false positive (liar).
An employee is chosen at random and given polygraph test,
says they are not a thief, but polygraph says liar.
What is prob that employee is lying?

$$\mathbb{P}(\text{Liar} \mid \text{Detected lie}) = \frac{\mathbb{P}(D|L)\mathbb{P}(L)}{\mathbb{P}(D|L)\mathbb{P}(L) + \mathbb{P}(D|L^c)\mathbb{P}(L^c)}$$

$$= \frac{(.95)(.01)}{(.95)(.01) + (.10)(.99)}$$

$$= .088$$ very low!

Expected value $\mathbb{E}[X] = \sum_{k \in \Omega} k \, \mathbb{P}(X=k)$

Conditional expectation $\mathbb{E}[Y \mid X=x] = \sum_y y \, \mathbb{P}(Y=y \mid X=x)$

Law of Total Expectation: Let Y be a random variable and $A_1, \cdots, A_k$ a partition of the state space Ω.

Then $\mathbb{E}[Y] = \sum_{i=1}^{k} \mathbb{E}[Y \mid A_i] \, \mathbb{P}(A_i)$

Example: Flip fair coin repeatedly. What is the expected # of flips to get 2 heads in a row?

Let Y = # flips

$\mathbb{E}[Y \mid 1^{st} \text{ flip is } T] = 1 + \mathbb{E}[Y]$ (1st flip no good, like starting over on 2nd flip)

$\mathbb{E}[Y \mid 1^{st} \text{ 2 flips } HT] = 2 + \mathbb{E}[Y]$ (1st 2 flips no good)

$\mathbb{E}[Y \mid 1^{st} \text{ 2 flips } HH] = 2$ (done!)

$\Rightarrow \mathbb{E}[Y] = (1 + \mathbb{E}[Y])(.5) + (2 + \mathbb{E}[Y])(.25) + 2(.25)$

Solve to get $\mathbb{E}[Y] = 6$

Law of Total Expectation $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]]$

Proof: $\mathbb{E}[\underbrace{\mathbb{E}[Y \mid X]}_{\text{treat as RV, depends on value } X \text{ takes}}] = \sum_x \mathbb{E}[Y \mid X = x]\,\mathbb{P}(X = x)$

$$= \sum_x \left(\sum_y y\,\mathbb{P}(Y = y \mid X = x) \right) \mathbb{P}(X = x)$$

$$= \sum_y y \left(\sum_x \mathbb{P}(Y = y \mid X = x)\,\mathbb{P}(X = x) \right)$$

(assuming interchange of sums is valid)

$$= \sum_y y\,\mathbb{P}(Y = y)$$

$$= \mathbb{E}[Y]$$

→ Example: Random sum of random variables

Let $T = X_1 + \cdots + X_N$, where N is also a RV & X_k iid RVs

iid = independent identically distributed

$$\mathbb{E}[T] = \mathbb{E}[\mathbb{E}[T \mid N]] = \mathbb{E}[N\,\mathbb{E}[X]]$$

$$= \mathbb{E}[N]\,\mathbb{E}[X] \quad \text{Wald's Eq}$$

Sections 2.1-2.2 Markov chains

Let Ω be a discrete set of states (can be finite or infinite)

A <u>Markov chain</u> is a sequence of RVs $X_0, X_1, X_2, \cdots$ such that

$$\mathbb{P}(X_{n+1} = j \mid X_0 = x_0, X_1 = x_1, \cdots, X_{n-1} = x_{n-1}, X_n = i)$$

$$= \mathbb{P}(X_{n+1} = j \mid X_n = i) \quad \text{for all states and } n \geq 0$$

That is, only X_n affects X_{n+1}, no other past values matter.

We can organize these probabilities into a <u>transition matrix</u> P

with entries $p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i)$

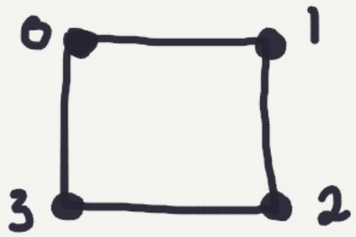
"one-step transition probabilities"

(probability of going from state i to j in one step)

Each row of P will sum to one, and all entries are nonnegative.

Such a matrix is called a "stochastic matrix."

Example: simple random walk on a graph



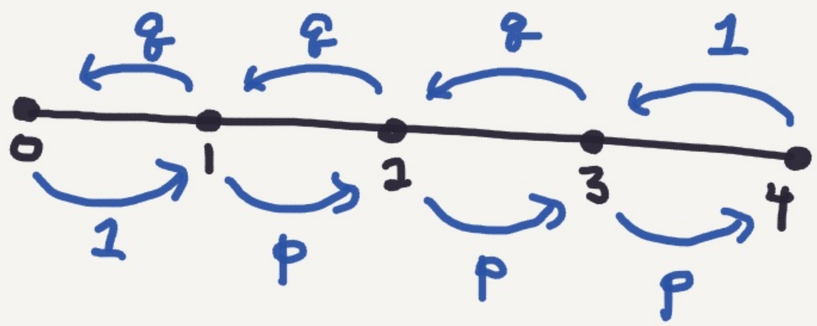
equally likely to follow any edge connected to current vertex

(for example, if currently at node 2, 50-50 chance of going to node 1 or 3)

Transition matrix

$$P = \begin{array}{c|c} & \text{to} \\ & \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \text{from} \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix} \end{array}$$

Example: Random walk with reflecting boundaries



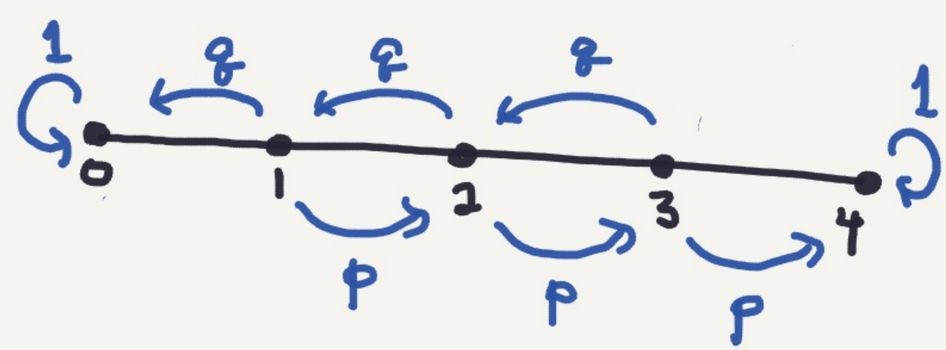
If at state 1, 2, or 3, flip a coin to determine whether to move right with probability p or left with probability $q = 1 - p$

If at state 0, go next to 1
If at state 4, go next to 3

Transition matrix

$$P = \begin{array}{c|c} & \text{to} \\ & \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \text{from} \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ 0 & q & 0 & p & 0 \\ 0 & 0 & q & 0 & p \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

R demo: Gambler's ruin
absorbing BCs



Probability vector: components nonnegative & sum to one
(gives probability of being in each state)

Initial prob vector: indicates how to start Markov chain
(prob distribution for X_0)

- can specify a particular state

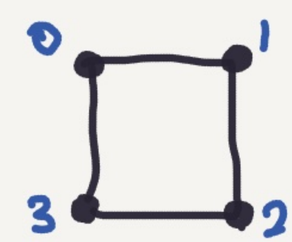
  $\alpha = [0 \ 0 \ 1 \ 0 \ 0]$

- can be a starting distribution with prob of beginning chain in each state

  $\alpha = [0 \ \tfrac{1}{2} \ 0 \ \tfrac{1}{2} \ 0]$

Multiplying α by P gives probability distribution for X_1

$$\begin{array}{c} \text{start at} \\ \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} \end{array} \begin{array}{c} \text{to} \\ \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix} \end{array} = \begin{array}{cccc} 0 & 1 & 2 & 3 \\ \end{array} \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$



Section 2.3

n-step transition probabilities

$$p_n(i,j) = \mathbb{P}(X_n = j \mid X_0 = i) = \mathbb{P}(X_{n+k} = j \mid X_k = i) \quad \text{for all } n, k \geq 0$$

(probability of going from state i to state j in exactly n steps)

$$\mathbb{P}(X_{n+1} = j \mid X_0 = i) = \sum_{k \in \Omega} \mathbb{P}(X_n = k \mid X_0 = i)\mathbb{P}(X_{n+1} = j \mid X_n = k)$$

all ways of getting to state j at time $n+1$

$$= \sum_k p_n(i,k)\, p(k,j)$$

matrix multiplication formula!

Using induction, we see $P_{n+1} = P_n P = P^n P = P^{n+1}$

where P_n has entries $p_n(i,j)$ (that is, $P_n = P^n$ for all $n \geq 0$)

If the initial prob vector is α, then the probability distribution for X_n is αP^n.

Chapman-Kolmogorov relationship

$$P_{m+n} = P_m P_n = P^m P^n \quad \text{(generalizes previous equation)}$$

Written out in terms of probabilities,

$$\mathbb{P}(X_{m+n} = j \mid X_0 = i) = \sum_k \mathbb{P}(X_n = k \mid X_0 = i)\, \mathbb{P}(X_{m+n} = j \mid X_n = k)$$

$$= \sum_k \mathbb{P}(X_n = k \mid X_0 = i)\, \mathbb{P}(X_m = j \mid X_0 = k)$$

by Markov property (transition probs are same at all steps)

$$= \sum_k p_n(i,k)\, p_m(k,j)$$

R demo: matrix powers

Example: Markov chain with state space $\{0,1\}$

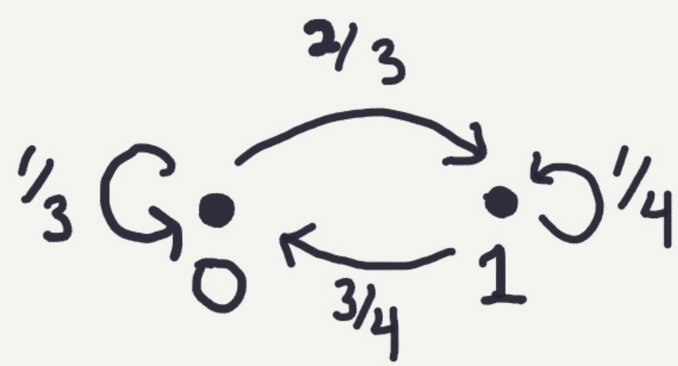
and transition matrix $P = \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 1/3 & 2/3 \\ 1 & 3/4 & 1/4 \end{array}$

Assume initial prob vector $\alpha = \begin{array}{cc} 0 & 1 \\ \hline 1 & 0 \end{array}$ (always start in state 0)

What is prob of being in state 1 at $n=3$?

$$\alpha P^3 = [1 \;\; 0] \begin{bmatrix} .495 & .505 \\ .568 & 432 \end{bmatrix} \quad \text{so } \mathbb{P}(X_3 = 1 \mid X_0 = 0) = .505$$

$$= \begin{array}{cc} 0 & 1 \\ \hline .495 & .505 \end{array}$$



Section 2.4 Long-term behavior and invariant prob.

What happens over the long run to $\mathbb{P}(X_n = j \mid X_0 = i)$ as $n \to \infty$?

We need to see what happens with P^n as $n \to \infty$.

If $\lim_{n\to\infty} \alpha P^n$ exists (converges to some π vector),

then we'll call π the "invariant prob vector"

(steady state, stationary, equilibrium vector)

In this case, does π depend on α?

R demo: 5 state reflecting boundaries – P^n does not converge

R demo: $P = \begin{bmatrix} 3/4 & 1/4 \\ 1/6 & 5/6 \end{bmatrix} \Rightarrow P^n$ converges to $\begin{bmatrix} .4 & .6 \\ .4 & .6 \end{bmatrix}$

$$\alpha P^n = [\alpha_0 \ \alpha_1] \begin{bmatrix} .4 & .6 \\ .4 & .6 \end{bmatrix} = [.4 \ .6]$$

Note $\alpha_0 + \alpha_1 = 1$

Same for all α!

Questions:

1) Do all stochastic matrices have at least one invariant prob vector (or at most one)?

2) What characterizes matrices with a <u>unique</u> invariant prob vector $\bar{\pi}$?

3) When is it true that $\lim_{n\to\infty} P^n = \begin{bmatrix} \bar{\pi} \\ \vdots \\ \bar{\pi} \end{bmatrix}$, which implies

$\lim_{n\to\infty} \alpha P^n = \bar{\pi}$ for all initial vectors α?

Some linear algebra will help us answer these questions.

Question 1: Transition matrix P has rows that sum to one,

so $$P\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix} = \begin{bmatrix}p_{11} & p_{12} & \cdots & p_{1n}\\ p_{21} & p_{22} & \cdots & p_{2n}\\ \vdots & & & \\ p_{m1} & p_{m2} & \cdots & p_{mn}\end{bmatrix}\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix} = \begin{bmatrix}\text{sum of row 1}\\ \text{sum of row 2}\\ \vdots\\ \text{sum of last row}\end{bmatrix} = \begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix}$$

(Eigenvalue λ with associated eigenvector $\vec{v}$ satisfies $P\vec{v} = \lambda\vec{v}$. A left eigenvector satisfies $\vec{v}P = \lambda\vec{v}$. Eigenvalues will be same, but left & right eigenvectors differ.)

Hence $\lambda = 1$ is an eigenvalue of P with right eigenvector $\begin{bmatrix}1\\\vdots\\1\end{bmatrix}$

P and P^t have the same eigenvalues, which implies

P has eigenvalue $\lambda = 1$ for some <u>left</u> eigenvector v : $vP = v$

which is an invariant prob vector.

Answer: every transition matrix P will have at least one invariant prob vector.

Question 2: When is there exactly one invariant prob vector?

R demo: absorbing boundaries example has multiple long-run behaviors, depending on initial state

note $\lambda = 1$ has multiplicity 2, and any linear combination of the associated left eigenvectors is a possible long-run distribution

Thm: If P is a stochastic matrix such that for some $n \geq 1$, P^n has all positive entries (that is, can reach all possible states from any initial vector), then P has eigenvalue $\lambda = 1$ with multiplicity 1, with unique associated left eigenvector π having all positive entries summing to one.

In addition, all other eigenvalues of P have abs value < 1 and $\lim_{n \to \infty} P^n = \begin{bmatrix} \pi \\ \vdots \\ \pi \end{bmatrix}$

The theorem also answered Question 3.

We now have two ways to determine $\bar{\pi}$ when P has $\lambda = 1$ with multiplicity 1

1. Raise matrix to high power to estimate $\lim_{n \to \infty} P^n$ and take any row as $\bar{\pi}$
2. Find the left eigenvector of P for $\lambda = 1$ (normalize so sums to one)

In R:

```
r <- eigen(t(P))
V <- r$vectors
lambda <- r$values
pibar <- V[,1]/sum(V[,1])
```

check which eigenvalue is 1, not always first one

Recap:

* If a stochastic process satisfies the property that the transition probabilities only depend on the current state and no past states, it is a Markov chain (assuming same transition matrix for all time steps).
* The one-step transition matrix P stores the transition prob of going from state i to state j in entry w/row i, column j.
* The n-step transition matrix is P^n, giving the prob of going from state i to state j in n steps.
* Given an initial prob vector α, the probability of ending up in state j after n steps is the jth component of αP^n.
* If P has exactly one eigenvalue $\lambda = 1$, then the left eigenvector normalized to sum to 1 gives the long-term prob for each state (proportion of time process is in that state), which agrees with the rows of $\lim_{n \to \infty} P^n$.

Chapter 3 Markov chains for the long term $\boxed{\text{R demo}}$

Section 3.1 Limiting distribution

Def'n: Let $X_0, X_1, X_2, \cdots$ be a Markov chain with transition matrix P. A <u>limiting distribution</u> for the Markov chain is a probability distribution $\bar{\lambda}$ with the property that for all states i and j,

$$\lim_{n\to\infty} P^n_{ij} = \lambda_j \quad (ij^{th} \text{ entry of } P^n \text{ equals } \lambda_j)$$

Equivalently, $\lim_{n\to\infty} \alpha P^n = \bar{\lambda}$ for all initial prob vectors α

$$\lim_{n\to\infty} P^n = \begin{bmatrix} \bar{\lambda} \\ \vdots \\ \bar{\lambda} \end{bmatrix} \quad (\text{all rows are copies of } \lambda)$$

Interpretation: λ_j gives proportion of time in state j over long run

$\lambda_j \approx$ prob $X_n = j$ for large n

Section 3.2 Stationary distribution

Def'n: Let $X_0, X_1, X_2, \cdots$ be a Markov chain with transition matrix P.
A <u>stationary distribution</u> is a prob distribution $\bar{\pi}$ that satisfies $\bar{\pi} = \bar{\pi} P$.

↳ also called <u>invariant</u> vector

That is, $\pi_j = \sum_i \pi_i P_{ij}$ for all states j.

Interpretation: If the current prob distr of states is $\bar{\pi}$, it will remain $\bar{\pi}$ for all future steps – each time we multiply by P, we get $\bar{\pi}$ back.

Suppose the initial prob vector is $\bar{\pi}$ (for X_0)
The prob vector for X_1 is then $\bar{\pi} P = \bar{\pi}$

X_2 $\quad \bar{\pi} P^2 = \bar{\pi} P = \bar{\pi}$

$\vdots$

X_n $\quad \bar{\pi} P^n = \bar{\pi}$

Question: How are limiting & stationary distributions related?

Answer: <u>Lemma 3.1</u> Limiting distributions are stationary distributions.

Proof: Suppose $\bar{\pi}$ is a limiting distribution:

$$\lim_{n\to\infty} (P^n)_{ij} = \pi_j \text{ for all states } i \text{ and } j$$

Need to show $\bar{\pi} = \bar{\pi} P$.

$$\pi_j = \lim_{n\to\infty} P^n_{ij} = \lim_{n\to\infty} P^{n+1}_{ij} = \lim_{n\to\infty} \sum_k \underbrace{P^n_{ik}}\, P_{kj}$$

$$= \sum_k \pi_k P_{kj}$$

$\Rightarrow \bar{\pi} = \bar{\pi} P$. $\square$

Converse is not true:

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$\bar{\pi} = [\,1/2 \;\; 1/2\,]$ is stationary but not limiting distr. because $\bar{\pi} = \bar{\pi} P$ yet $P^{odd} = P$ & $P^{even} = I$.

Matrices with all positive entries are particularly nice – can get to any state from any state with some positive probability.

Def'n: A transition matrix P of a Markov chain is called <u>regular</u> if some power of P has all positive entries ("$P^n > 0$" for some $n \geq 1$)

Example: $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ not regular because

$P^{odd} = P$ & $P^{even} = I$ have zeros

Example: $P = \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}$

$$P^2 = \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix}$$

yes, regular

Theorem 3.2 Limit Theorem for Regular Markov Chains

Let P be the transition matrix for a Markov chain.

Suppose P is regular (P^n has all positive entries for some $n \geq 1$).

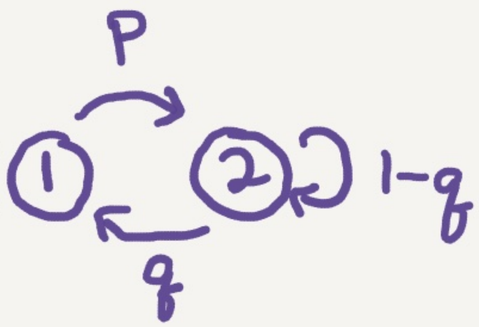
Then P has a limiting distribution which is the unique, positive, stationary distribution $\bar{\pi}$ of the Markov chain: $\bar{\pi} = \bar{\pi} P$

(invariant)

all states get visited

$$\lim_{n \to \infty} P^n = \begin{bmatrix} \bar{\pi} \\ \vdots \\ \bar{\pi} \end{bmatrix}$$

Example: 2-state chain



Transition matrix $P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$ is regular if $p, q > 0$

Solve $\bar{\pi} = \bar{\pi} P$: $[\pi_1 \ \pi_2] = [\pi_1 \ \pi_2] \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$

$$= [\pi_1(1-p) + \pi_2 q \quad \pi_1 p + \pi_2(1-q)]$$

Solve $\pi_1 = \pi_1(1-p) + \pi_2 q$
$\pi_2 = \pi_1 p + \pi_2(1-q)$ $\rightarrow$ $p\pi_1 - q\pi_2 = 0$
$p\pi_1 - q\pi_2 = 0$ $\rightarrow$ $\pi_2 = \frac{p}{q}\pi_1$

Need $\pi_1 + \pi_2 = 1$

$$\Rightarrow \bar{\pi} = \left[\frac{q}{p+q} \quad \frac{p}{p+q} \right]$$

Section 3.3 Classification of states

"Can you find the way to state a?"

In general, we can't expect to have a unique (stationary) invariant prob vector $\bar{\pi}$.

Example: a simple chain with reflecting boundaries has different limits for P^n with n odd vs n even.

Example: a chain with absorbing states has different limits corresponding to each absorbing state (start there, stay there)

Example: has two components, can never get from one to the other, so different limiting distr depending where chain starts

We want to classify these different situations.

Def'n: We say states i & j of a Markov chain <u>communicate</u> with each other ("$i \leftrightarrow j$") if there exists $m, n \geq 0$ such that $P^m_{ij} > 0$ and $P^n_{ji} > 0$. That is, the chain has a positive prob of going from state i to state j in a finite # of steps, and vice versa.

This defines an equivalence relation on the state space

1) Reflexive: $i \leftrightarrow i$ (in 0 steps)
2) Symmetric: $i \leftrightarrow j$ if and only if $j \leftrightarrow i$ (by the def'n)
3) Transitive: if $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$

Equivalence relations partition the space into disjoint sets, which here we will call <u>communication classes</u>.

Def'n: A Markov chain is called <u>irreducible</u> if it has only one communication class, that is, there is a positive prob of getting from every state to every other state in a finite # of steps.

Otherwise, chain is called <u>reducible</u>.

If a chain leaves a communication class eventually with probability 1, then states in that class are called <u>transient</u>.
Otherwise, called <u>recurrent</u> (never leaves that class).

Example: Chain w/ reflecting boundaries

irreducible
$\{1, 2, 3, 4\}$
recurrent

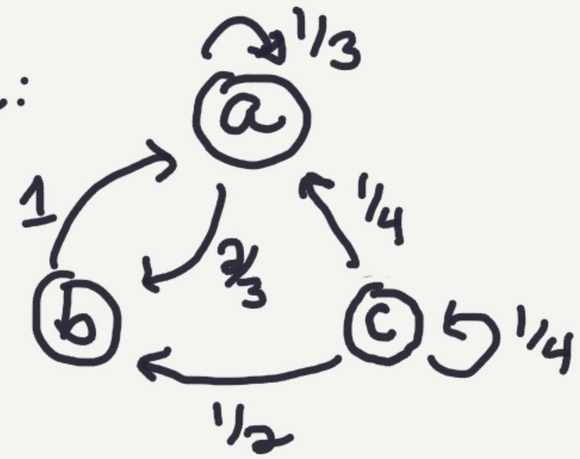
Chain w/ absorbing boundaries

Communication classes: $\{1\}$ $\{5\}$ $\{2, 3, 4\}$

recurrent states

transient

Example:



Communication classes: $\{c\}$ transient

$\{a, b\}$ recurrent

Questions: On average, how many steps until leave state c forever?

How many steps on average to return to a from a, or get to b from a?

Defn: The <u>first passage time</u> to state j is $T_j = \min\{n>0 : X_n = j\}$

(if $X_n \neq j$ for all $n>0$, say $T_j = \infty$)

The prob of ever returning to j is $f_j = \mathbb{P}\{T_j < \infty : X_0 = j\}$

For recurrent states, $f_j = 1$; for transient states, $f_j < 1$.

In example above, $f_a = 1$ $\quad f_b = 1$ $\quad f_c = 1/4$

Whether a state is eventually revisited is related to how often we expect that state to be visited.

Start chain at state $X_0 = i$ and let j be another state.

Let $I_n = \begin{cases} 1 & X_n = j \\ 0 & X_n \neq j \end{cases}$, so $\sum_{n=0}^{\infty} I_n$ = # visits to j

"indicator fn"

$$\mathbb{E}\left[\sum_{n=0}^{\infty} I_n\right] = \sum_{n=0}^{\infty} \mathbb{E}[I_n] = \sum_{n=0}^{\infty} \mathbb{P}\{X_n = j \mid X_0 = i\} = \sum_{n=0}^{\infty} P^n_{ij}$$ (may or may not converge)

(expected # of visits to state j)

Suppose state j is <u>recurrent</u>, so chain started at j eventually revisits j.
Once chain hits j, process begins anew behaving like new chain started at j.
Hence we expect j to be visited infinitely often: $\sum_{n=0}^{\infty} P^n_{jj} = \infty$ diverges.

Now suppose state j is <u>transient</u>.

Prob of hitting j equals f_j and of never hitting f_j is $1-f_j$.

Think of situation as a series of coin flips : $\mathbb{P}\{\text{head}\} = f_j$

$\mathbb{P}\{\text{tail}\} = 1-f_j$

Flip until tail to see how often visit j before leaving forever.

$$\mathbb{E}[\text{flips until tail}] = \underset{\uparrow\ 1^{st}\text{ flip H}}{f_j}\,(\mathbb{E}[\text{flips until tail}]+1) + \underset{\uparrow\ 1^{st}\text{ flip T}}{(1-f_j)}\cdot 1$$

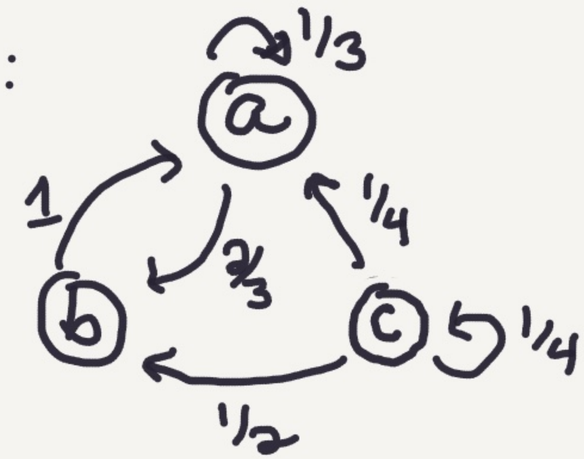
(condition on first flip)

$$= f_j\,\mathbb{E}[\text{flips until tail}] + 1$$

$$\Rightarrow \mathbb{E}[\text{flips until tail}] = \frac{1}{1-f_j}$$

Hence expected # visits to state j is $\sum_{n=0}^{\infty} P^n_{jj} = \frac{1}{1-f_j}$

Return to example:



$$P = \begin{array}{c|ccc} & \text{to } a & b & c \\ \hline \text{from } a & 1/3 & 2/3 & 0 \\ b & 1 & 0 & 0 \\ c & 1/4 & 1/2 & 1/4 \end{array}$$

State a: $P^0_{aa} = 1$

$P^1_{aa} = 1/3$

$P^2_{aa} = 2/3$

$$P^n_{aa} = \frac{1}{3} P^{n-1}_{aa} + \frac{2}{3} P^{n-2}_{aa}$$

a to a in n-1 steps
a to a in n^{th} step

a to a in n-2 steps
a to b then b to a

$\geq 1/3$ by induction

$$\Rightarrow \sum_{n=0}^{\infty} P^n_{aa} = \infty$$

State c: $\sum_{n=0}^{\infty} P^n_{cc} = \sum_{n=0}^{\infty} (1/4)^n = \frac{1}{1-1/4} = \frac{4}{3} = \frac{1}{1-f_c}$, so $f_c = 1/4$

(have to return to c every step)

geometric series formula $\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}$ for $|r|<1$

Theorem on Recurrence and Transience

For a Markov chain with a finite state space and transition matrix P,

$$\sum_{n=0}^{\infty} P_{jj}^{n} = \infty \iff \text{state } j \text{ is recurrent}$$

$$\sum_{n=0}^{\infty} P_{jj}^{n} < \infty \iff \text{state } j \text{ is transient}$$

<u>Corollary</u>: For finite-state, irreducible Markov chains, all states are recurrent.

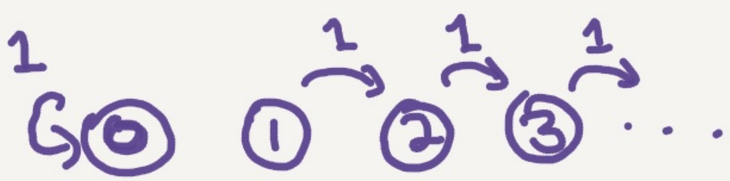
(See proof of Thm 3.3, which shows $\sum_n P_{ij}^n$ converges $\iff$ $\sum_n P_{jj}^n$ converges whenever states i & j are in same communication class)

Defn: A set C of states is called <u>closed</u> if no state outside of C is accessible from any state inside C.

If C is closed, then $P_{ij} = 0$ for all $i \in C$, $j \notin C$.

<u>Lemma 3.5</u>: A communication class with only recurrent states is closed.

A <u>finite</u> communication class is closed <u>only if</u> it consists of recurrent states.

Example: 

(why infinite chains are tricky)

$\{0\}$ recurrent & closed

$\{1, 2, 3, \cdots\}$ transient & closed

<u>Canonical decomposition</u> : partition a finite Markov chain into communication classes R_i of recurrent states and a set T of transient states.

Reorder states to organize transition matrix as

$$P = \begin{array}{c} \\ T \\ R_1 \\ \vdots \\ R_m \end{array} \begin{array}{c} \begin{array}{cccc} T & R_1 & \cdots & R_m \end{array} \\ \begin{bmatrix} * & * & \cdots & * \\ 0 & [\ ] & & 0 \\ \vdots & 0 & & \vdots \\ 0 & 0 & & [\ ] \end{bmatrix} \end{array}$$

square blocks for each R_i
taking powers of P gives
same structure with blocks $[\ ]^n$

This breaks down the Markov chain into components that are easier to analyze for long term behavior – the chain eventually leaves the transient states and remains in one of the recurrent classes.

Section 3.4 Irreducible Markov chains

Recall that a Markov chain is called <u>irreducible</u> if it has only one communication class, that is, there is a positive prob of getting from every state to every other state in a finite # of steps.

<u>Theorem 3.6</u> Limit Theorem for Finite Irreducible Markov chains

Suppose $X_0, X_1, X_2, \cdots$ is a <u>finite</u> <u>irreducible</u> Markov chain.

For each state j, let $\mu_j = \mathbb{E}[T_j \mid X_0 = j]$ expected return time to state j

Then μ_j is finite and there exists a unique stationary distribution $\bar{\pi}$ such that $\pi_j = \frac{1}{\mu_j}$ for all states j.

Furthermore, $\pi_j = \lim_{n\to\infty}\left(\frac{1}{n}\sum_{m=0}^{n-1} P_{ij}^{m}\right)$ for all states i and j

any initial state

prob of going from i to j in m steps

(See rigorous proof in section 3.10 – we'll just get some intuition on why this is true.)

<u>Intuition</u>: π_j = proportion time spent in state j

Given a finite, irreducible MC, start at state j $(X_0 = j)$ and let $T_j = \min\{n \geq 1 : X_n = j\}$ first return time

$\mathbb{E}[T_j \mid X_0 = j] < \infty$ because finite # of states and positive prob of getting between any pair of steps in a finite time

Consider k^{th} return to state j: $T_j^{(1)} + T_j^{(2)} + \cdots + T_j^{(k)}$

(all independent identically distributed RVs)

Law of Large Numbers says $\frac{1}{k}\left(T_j^{(1)} + T_j^{(2)} + \cdots + T_j^{(k)}\right) \approx \mathbb{E}[T_j]$

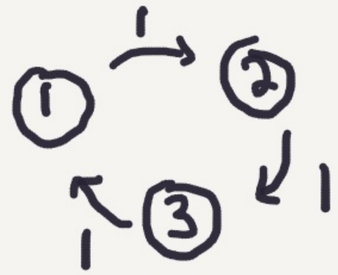
(sample mean $\to$ true mean as sample size $\to \infty$)

So we expect around k visits to state j in $n = k\,\mathbb{E}[T_j]$ steps

We also expect around $n\pi_j$ visits to j after n steps,

so $\underset{\text{visits}}{k} \approx \underbrace{k\,\mathbb{E}[T_j]}_{n \text{ steps}}\pi_j \implies \boxed{\mathbb{E}[T_j] = \frac{1}{\pi_j}}$

Example:



$\pi_1 = 1/3$

$\mathbb{E}[T_1] = 3$

$\boxed{\text{R demo}}$

$$P = \begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & .4 & .2 & .4 \\ 2 & .6 & 0 & .4 \\ 3 & .2 & .5 & .3 \end{array} \qquad \begin{array}{ccc} \pi_1 & \pi_2 & \pi_3 \\ \end{array}$$

$$\bar{\pi} = [\,.3\overline{78} \quad .2\overline{57} \quad .3\overline{6}\,]$$

Start in state 2 $\Rightarrow$ $\mathbb{E}[T_2] = \frac{1}{\pi_2} \approx 3.88$

Simulation shows spread of return times

For <u>infinite</u> state spaces, can have more complicated situations.

Def'n: State j is called <u>positive recurrent</u> if Markov chain will eventually revisit state j with prob 1 and expected first return time is <u>finite</u>.

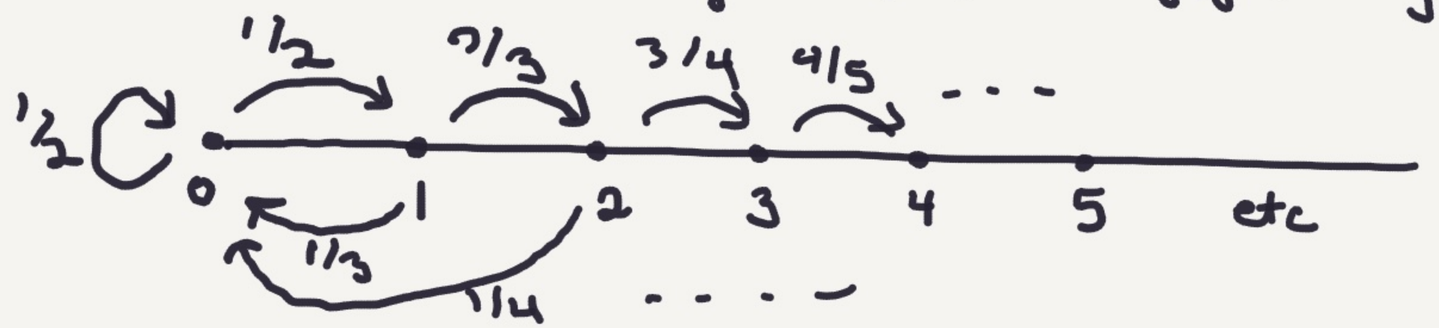
State j is called <u>null recurrent</u> if Markov chain will eventually revisit state j with prob 1 but expected first return time is <u>infinite</u>. (can only occur for infinite state space)

State j is called <u>transient</u> if the prob of returning to j is less than 1.

All states in the same communication class will share the same classification.

Example: State space $\{0, 1, 2, \cdots\}$

Transition prob $P_{j0} = \frac{1}{j+2}$ $\quad P_{j,j+1} = \frac{j+1}{j+2}$ for $j \geq 0$



First return time to state 0: look at prob of <u>never</u> returning

$$\mathbb{P}\{T_0 = \infty \mid X_0 = 0\} = \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} \cdots$$

$$= \prod_{j=0}^{\infty} \frac{j+1}{j+2} = \lim_{j \to \infty} \frac{1}{j+2} = 0$$

So the prob of returning to state 0 is 1.

The prob of returning to state 0 in exactly k steps is

$$\mathbb{P}\{T_0 = k \mid X_0 = 0\} = \underbrace{\frac{1}{2} \cdot \frac{2}{3} \cdots \frac{k-1}{k}}_{\text{march out to state } j+1 = k-1} \cdot \underbrace{\frac{1}{(k-1)+2}}_{\text{return to } 0} = \frac{1}{k(k+1)}$$

Expected first return time to state 0: $\mathbb{E}[T_0 \mid X_0 = 0] = \sum_{k=1}^{\infty} k \mathbb{P}\{T_0 = k \mid X_0 = 0\}$

$$= \sum_{k=1}^{\infty} \frac{k}{k(k+1)} = \sum_{k=1}^{\infty} \frac{1}{k+1}$$

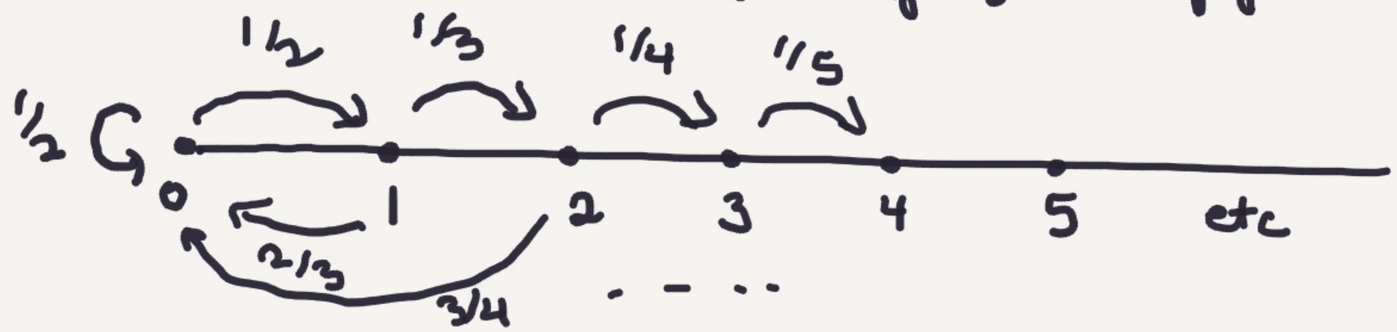
Harmonic series, diverges

Hence state 0 is null recurrent.

Irreducible MC, so all states are <u>null recurrent.</u>

Example: State space $\{0,1,2,\cdots\}$

Transition prob $P_{j0} = \frac{j+1}{j+2}$ $\quad P_{j,j+1} = \frac{1}{j+2}$



$$\mathbb{P}\{T_0 = \infty \mid X_0 = 0\} = \frac{1}{2}\cdot\frac{1}{3}\cdot\frac{1}{4}\cdots = \prod_{j=0}^{\infty}\frac{1}{j+2} = 0$$

so returns to 0 with prob 1

$$\mathbb{P}\{T_0 = k \mid X_0 = 0\} = \frac{1}{2}\cdot\frac{1}{3}\cdots\underbrace{\frac{1}{(k-2)+2}}_{\text{march out to state } j+1=k-1}\cdot\underbrace{\frac{k}{k+1}}_{\text{return to 0 from } j=k-1} = \frac{k}{(k+1)!}$$

Expected first return time $\mathbb{E}[T_0 \mid X_0 = 0] = \sum_{k=1}^{\infty} k\cdot\frac{k}{(k+1)!} < \infty$

by ratio test

Hence state 0 is positive recurrent,

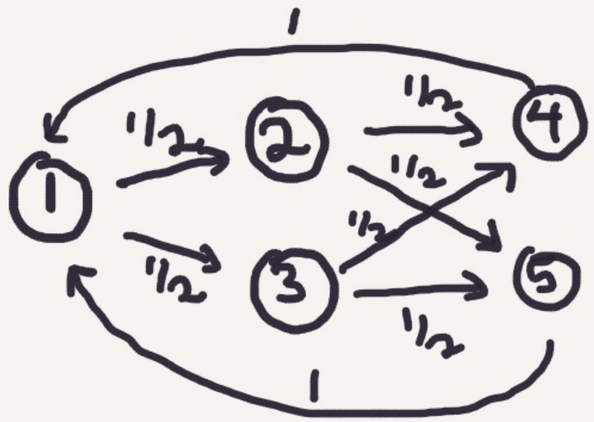
and so all states are <u>positive recurrent</u>

(irreducible MC, single communication class)

Section 3.5 Periodicity (in a stochastic sense)

Def'n: The <u>period</u> $d = d(i)$ of a state i of a MC is the greatest common divisor of $J_i = \{n \geq 0 : P^n_{ii} > 0\}$

Example:



$\boxed{\text{R demo}}$

P^{3m} same for $m \geq 1$, all positive diagonal entries

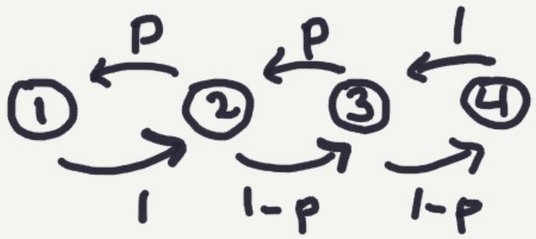
$J_i = \{0, 3, 6, 9, \cdots\}$ for all states i

Period $d = d(i) = 3$ for all i

Lemma 3.7 All states of an irreducible MC have the same period. (See proof in book.)

Def'n: We call an irreducible MC <u>aperiodic</u> if period equals 1.

Example: 4-state reflecting boundaries



$\boxed{\text{R demo}}$

P^{2m} has positive diagonal entries for $m \geq 1$ but odd powers of P have zeros

$J_i = \{0, 2, 4, 6, \cdots\}$ and $d = 2$ for all states

Theorem: If P is the transition matrix for an irreducible, finite, aperiodic Markov chain, then there exists a unique stationary vector $\bar{\pi}$ satisfying $\bar{\pi} P = \bar{\pi}$. If $\bar{\alpha}$ is any initial prob vector, then $\lim_{n\to\infty} \bar{\alpha} P^n = \bar{\pi}$. Moreover, $\pi_i > 0$ for all states i.

visit every state over long run

initial state doesn't matter in long run

Proof: We can apply our prior thm on irreducible MCs if we can show $P^n > 0$ for some $n \geq 1$ (P is regular).

Take any states i and j. Since P is irreducible, $P_{ij}^{m_{ij}} > 0$ for some $m_{ij} \geq 1$

Because P is aperiodic, there exists M_i such that for all $n \geq M_i$, $P_{ii}^n > 0$

(number theory result)

Hence for all $n \geq M_i$, $P_{ij}^{n+m_{ij}} \geq P_{ii}^n P_{ij}^{m_{ij}} > 0$

(one way to get from i to j in $n + m_{ij}$ steps)

Let $M = \max\{M_i + m_{ij} : \text{all states } i, j\}$

Then $P_{ij}^n > 0$ for all $n \geq M$ and states i, j.

That is, $P^M > 0$, so the prior theorem applies. $\square$

What about reducible & periodic chains?. What can we say?

Suppose P is reducible with recurrent classes $R_1, \cdots, R_r$
(once in one of these classes, stays there)

and transient classes $T_1, \cdots, T_s$
(eventually leaves)

If the recurrent classes are aperiodic (treated as their own MC), then we have a stationary vector π^k for each R^k
(can set $\pi^k_i = 0$ for $i \notin R^k$)

i^{th} component of vector π^k

Then $\lim_{n\to\infty} P^n_{ij} = \begin{cases} \pi^k_j & \text{for } i,j \in R^k \\ 0 & \text{o.w.} \end{cases}$

$\lim_{n\to\infty} P^n_{ij} = 0$ for all transient states

Note that if we have more than one recurrent class, where the chain starts will be important for which recurrent class it eventually ends up in.

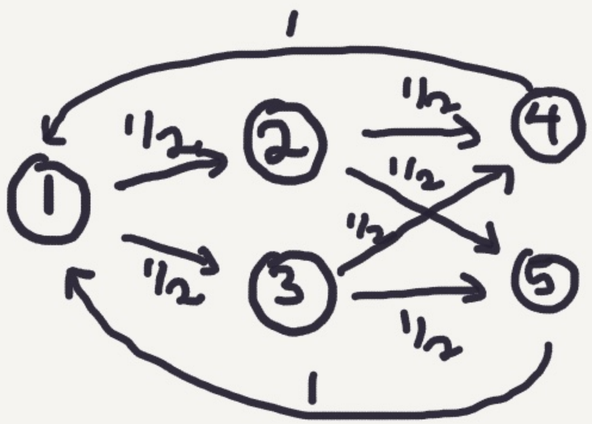
Now suppose the Markov chain is irreducible but periodic, $d > 1$

In this case, the chain splits into d sets of states $A_1, \cdots, A_d$

such that

$$\begin{tikzcd} A_1 \arrow[r] & A_2 \arrow[dr] & \\ A_d \arrow[u] & \cdots \arrow[l] & A_3 \arrow[l] \end{tikzcd}$$

Consider previous example:



$\boxed{\text{R demo}}$

Eigenvalues of P are

$$\lambda = \underbrace{1, \ -\tfrac{1}{2} \pm \tfrac{\sqrt{3}}{2} i}_{3^{rd} \text{ roots of unity}}, 0, 0$$

Eigenvector for $\lambda = 1$ is $[1/3 \ \ 1/6 \ \ 1/6 \ \ 1/6 \ \ 1/6]$

A bit of linear algebra will reveal what's happening for period d chains.

Most transition matrices are "diagonalizable" (except for rare cases)

The idea is that an $n \times n$ matrix P has n eigenvalues (can be duplication) associated with n linearly independent eigenvectors.

Let $D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$ diagonal matrix of eigenvalues

Q = matrix with associated right eigenvectors $P\bar{w}_k = \lambda_k \bar{w}_k$

$= [\bar{w}_1 \ \bar{w}_2 \ \cdots \ \bar{w}_n]$

Then $Q^{-1}PQ = Q^{-1}[P\bar{w}_1 \cdots P\bar{w}_n] = Q^{-1}[\lambda_1\bar{w}_1 \cdots \lambda_n\bar{w}_n]$

$$= Q^{-1}\underbrace{[\bar{w}_1 \cdots \bar{w}_n]}_{Q}\underbrace{\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}}_{D}$$

$$= D$$

Hence $D = Q^{-1}PQ$,

so $DQ^{-1} = Q^{-1}P$

Let $\bar{v}_1, \cdots, \bar{v}_n$ be the rows of $Q^{-1} = \begin{bmatrix} \bar{v}_1 \\ \vdots \\ \bar{v}_n \end{bmatrix}$

$$DQ^{-1} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}\begin{bmatrix} \bar{v}_1 \\ \vdots \\ \bar{v}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\bar{v}_1 \\ \vdots \\ \lambda_n\bar{v}_n \end{bmatrix} \Rightarrow \bar{v}_k P = \lambda_k P, \ k = 1, \cdots, n$$

$$\| $$

$$Q^{-1}P = \begin{bmatrix} \bar{v}_1 \\ \vdots \\ \bar{v}_n \end{bmatrix} P = \begin{bmatrix} \bar{v}_1 P \\ \vdots \\ \bar{v}_n P \end{bmatrix}$$

Rows of Q^{-1} are the left eigenvectors of P.

One of the nice properties of diagonalization is taking powers of the matrix:

$$D = Q^{-1}PQ \Rightarrow P = QDQ^{-1} \Rightarrow P^2 = (QDQ^{-1})(QDQ^{-1})$$
$$= QD^2Q^{-1}$$
$$= \begin{bmatrix} \bar{w}_1 & \cdots & \bar{w}_n \end{bmatrix} \begin{bmatrix} \lambda_1^2 & & \\ & \ddots & \\ & & \lambda_n^2 \end{bmatrix} \begin{bmatrix} \bar{v}_1 \\ \vdots \\ \bar{v}_n \end{bmatrix}$$

Generally, $P^m = QD^mQ^{-1}$

$$= \begin{bmatrix} \bar{w}_1 & \cdots & \bar{w}_n \end{bmatrix} \begin{bmatrix} \lambda_1^m & & \\ & \ddots & \\ & & \lambda_n^m \end{bmatrix} \begin{bmatrix} \bar{v}_1 \\ \vdots \\ \bar{v}_n \end{bmatrix}$$
$$= \lambda_1^m \bar{w}_1 \bar{v}_1 + \cdots + \lambda_n^m \bar{w}_n \bar{v}_n$$

If we knew more about $\lambda_1, \cdots, \lambda_n$, we could see what happens as the power m grows large (long term behavior).

Perron-Frobenius Theorem for Irreducible Periodic MCs

Let P be the transition matrix for a finite, irreducible MC with period d. Then the following hold:

1) P has a single eigenvalue $\lambda = 1$ with right eigenvector $\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ and left eigenvector with all positive components.

2) P has d eigenvalues that are d^{th} roots of unity:

   (Let $\omega = e^{2\pi i/d} = \cos(2\pi/d) + i\sin(2\pi/d)$, so $\omega^d = 1$ & $|\omega| = 1$
   then d^{th} roots of unity are $1, \omega, \omega^2, \dots, \omega^{d-1}$)

3) Remaining eigenvalues have absolute value less than 1.

   $\hookrightarrow \lambda_{d+1}^m, \dots, \lambda_n^m \to 0$ as $m \to \infty$

This means that $P^m \approx \begin{bmatrix} \bar{v}_1 \\ \vdots \\ \bar{v}_1 \end{bmatrix} + \omega^m \bar{w}_2 \bar{v}_2 + \omega^{2m} \bar{w}_3 \bar{v}_3 + \dots + \omega^{(d-1)m} \bar{w}_{d-1} \bar{v}_{d-1}$

(why we get d alternating matrices for large powers)

We can also think about the average time spent in each state by averaging across the cycle: Proportion time spent in state j if start at state i $= \lim_{m\to\infty} (P_{ij}^{m+1} + \dots + P_{ij}^{m+d})/d$

R demo

Section 3.6 Ergodic Markov Chains

Def'n: A Markov chain is called <u>ergodic</u> if it is

1) irreducible
2) aperiodic
3) all states have finite expected return times

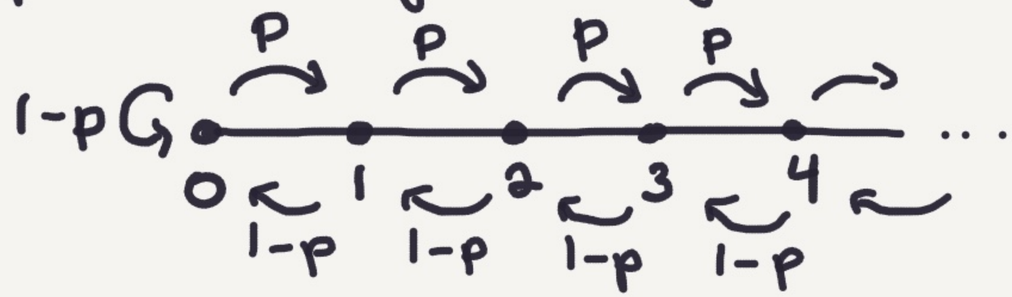
<u>Theorem 3.8</u> Fundamental Limit Theorem for Ergodic Markov Chains

Let $X_0, X_1, X_2, \cdots$ be an ergodic MC. Then there exists a unique stationary distribution $\bar{\pi}$ with all positive components that is the limiting distribution of the chain: $\pi_j = \lim_{n\to\infty} P_{ij}^n$ for all states i, j

$\hookrightarrow (\bar{\pi} P = \bar{\pi})$

(See proof in section 3.10. Essentially need to prove ergodic MC $\Leftrightarrow$ P is regular.)

Example: Partially reflecting walk on infinite state space



Ergodic?

Find the stationary distribution: $\sum_{j=0}^{\infty} \pi_j = 1$ and $\underbrace{\sum_{i=0}^{\infty} \pi_i P_{ij} = \pi_j \text{ for all } j}$

$\pi_1(1-p) + \pi_0(1-p) = \pi_0$

$\pi_{j+1}(1-p) + \pi_{j-1} p = \pi_j, \; j>0$

Case $p = 1/2$: $\pi_1 = \pi_0, \; \pi_2 = \pi_0, \ldots$

Can't get $\sum \pi_j = 1$, so no sol'n

<u>Not</u> ergodic. (In fact, expected return times are infinite and states are null recurrent.)

Case $p \neq 1/2$: Assume sol'n of form $\pi_j = x^j$:

find constants that work

$$x^{j+1}(1-p) + x^{j-1} p = x^j$$

$$\Rightarrow x^2(1-p) - x + p = 0$$

$$\Rightarrow x = \frac{1 \pm \sqrt{(1-2p)^2}}{2(1-p)}$$

$$= 1, \; p/1-p$$

Possible sol'ns are $\pi_j = c_1 + c_2 \left(\frac{p}{1-p}\right)^j$

If $p > 1/2$, then $\frac{p}{1-p} > 1$ and no sol'n exists with $\sum \pi_j = 1$. <u>Not</u> ergodic. (can show all states are transient)

If $p < 1/2$, let $c_1 = 0$, $c_2 = 1/\sum_{j=0}^{\infty} \left(\frac{p}{1-p}\right)^j = \frac{1-2p}{1-p}$

to get $\pi_j = \frac{1-2p}{1-p}\left(\frac{p}{1-p}\right)^j$ (can show is ergodic and all states are positive recurrent)

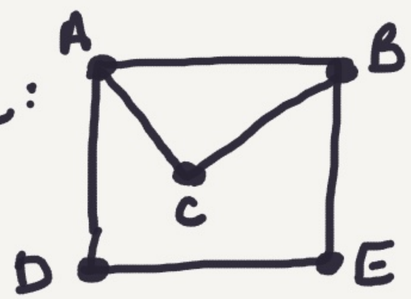
Section 3.7 Reversibility

Can you tell whether the chain is running forward or backward?.

Defn: An irreducible MC with transition matrix P and stationary distr. π ($\pi P = \pi$)

is <u>time-reversible</u> if $\pi_i P_{ij} = \pi_j P_{ji}$ for all states i & j

Example:



Equally likely to follow any edge

$$P = \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 1/2 & 0 & 1/2 & 0 \end{bmatrix} \end{array}$$

Reversible?. $\pi = [\, 3/12 \;\; 3/12 \;\; 2/12 \;\; 2/12 \;\; 2/12 \,]$ (degree of vertex divided by sum of all degrees)

If i & j are neighbors, $\pi_i P_{ij} = \dfrac{\#\text{edges}(i)}{\sum_k \#\text{edges}(k)} \cdot \dfrac{1}{\#\text{edges}(i)} = \dfrac{1}{2(\#\text{edges})}$

(state i has $\#\text{edges}(i)$ neighbors)

total in graph

$\pi_j P_{ji}$ also equals $\dfrac{1}{2(\#\text{edges})}$

Yes, reversible.

Prop 3.9 Let P be the transition matrix of an irreducible MC.

If $\bar{x}$ is a prob distribution vector with $x_i P_{ij} = x_j P_{ji}$ for all i, j, then $\bar{x}$ is the stationary distr $\bar{\pi}$ and the MC is reversible.

Proof: Need to show $\bar{x} = \bar{x}P$. Consider j^{th} component:

$$(\bar{x}P)_j = \sum_i x_i P_{ij}$$

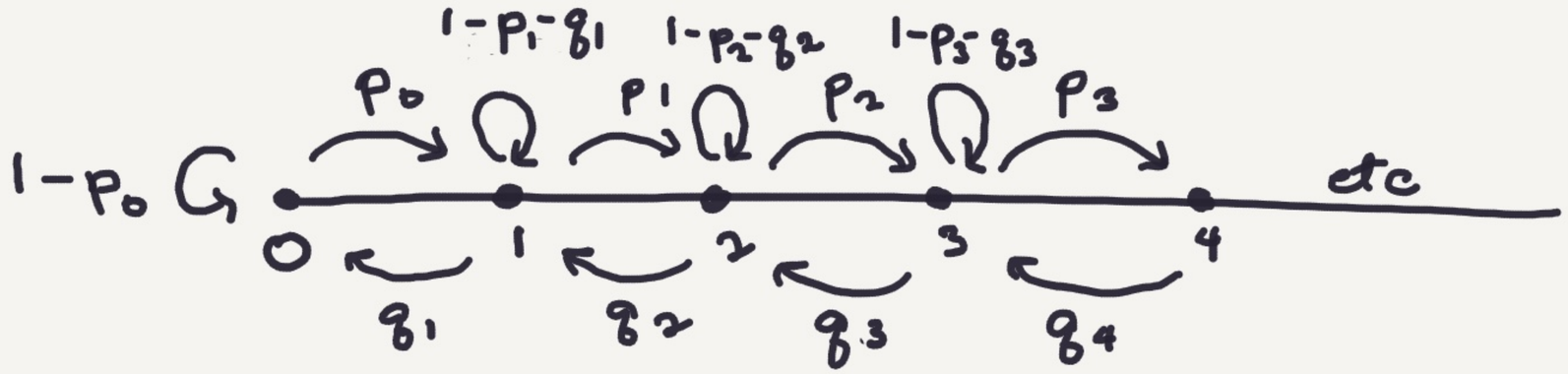
$$= \sum_i x_j P_{ji} \quad \text{using property given for } \bar{x}$$

$$= x_j \cancel{\sum_i P_{ji}}^{\,1} \quad \text{(summing across } j^{th} \text{ row of } P\text{)}$$

$$= x_j$$

So $\bar{x} = \bar{x}P$.

Example: birth and death chain



Solve for π_j satisfying $\pi_i P_{ij} = \pi_j P_{ji}$ to find when MC is reversible.

Only need to check moving over one spot:

$$\underbrace{\pi_i P_{i,i-1}}_{\text{move left}} = \underbrace{\pi_{i-1} P_{i-1,i}}_{\text{move right}} \longleftrightarrow \pi_i q_i = \pi_{i-1} p_{i-1} \quad \text{for } i \geq 1$$

So want $\pi_1 q_1 = \pi_0 p_0 \rightarrow \pi_1 = \frac{p_0}{q_1} \pi_0$

$$\pi_2 q_2 = \pi_1 p_1 \rightarrow \pi_2 = \frac{p_1}{q_2} \pi_1 = \frac{p_1}{q_2} \cdot \frac{p_0}{q_1} \pi_0$$

$$\vdots$$

$$\pi_k q_k = \pi_{k-1} p_{k-1} \rightarrow \pi_k = \frac{p_{k-1}}{q_k} \pi_{k-1} = \frac{p_{k-1} p_{k-2} \cdots p_1 p_0}{q_k q_{k-1} \cdots q_1} \pi_0$$

To get $\sum_{k=0}^{\infty} \pi_k = 1$, let $\pi_0 = 1 \Big/ 1 + \sum_{k=1}^{\infty} \prod_{i=1}^{k} \frac{p_{i-1}}{q_i}$, assuming series converges.

Birth&death chain being reversible requires that convergence.

Note: if all $p_i = p$ & $q_i = q$, then $\pi_k = \pi_0 \left(\frac{p}{q}\right)^k$ and $\pi_0 = 1 \Big/ \sum_{k=0}^{\infty} \left(\frac{p}{q}\right)^k = \frac{1}{1/1-\frac{p}{q}} = 1 - \frac{p}{q}$

Section 3.8 Absorbing chains

Idea: Decompose state space into recurrent classes $R_1, \cdots, R_m$ and a set of transient states. Treat each recurrent class as an absorbing state to calculate how long on average it takes for the MC to enter R_j, the prob of entering that particular R_j, how many visits to transient states the MC made before being absorbed into R_j, etc.

Reorganize transition matrix

$$P = \begin{array}{c} \\ T \\ R_1 \\ \vdots \\ R_m \end{array} \begin{array}{c} \begin{array}{cc} T & R_1 \cdots R_m \end{array} \\ \left[\begin{array}{c|c} Q & R \\ \hline 0 & \underline{1} \end{array}\right] \end{array}$$

treat each R_j as a single combined state

OR

$$P = \begin{array}{c} \\ T \\ R_1 \\ \vdots \\ R_m \end{array} \begin{array}{c} \begin{array}{cc} T & R_1 \cdots R_m \end{array} \\ \left[\begin{array}{c|c} Q & R \\ \hline 0 & \tilde{P} \end{array}\right] \end{array}$$

if keeping all states

$\tilde{P}$ is a stochastic matrix (rows sum to 1)

Q has rows that sum to ≤ 1

Block matrix multiplication

$$\begin{array}{c|cc} & T & R_1 \cdots R_m \\ \hline T & Q & R \\ R_1 \vdots R_m & 0 & \tilde{P} \end{array} \begin{bmatrix} Q & R \\ 0 & \tilde{P} \end{bmatrix} = \begin{bmatrix} Q^2 & QR + R\tilde{P} \\ 0 & \tilde{P}^2 \end{bmatrix}$$

$$\begin{bmatrix} Q & R \\ 0 & \tilde{P} \end{bmatrix}^n = \begin{bmatrix} Q^n & * \\ 0 & \tilde{P}^n \end{bmatrix}$$

States in T are transient so $Q^n \to 0$ as $n \to \infty$

This implies no eigenvalue of Q can equal 1 $\quad (\lambda^n \to 0$ as $n \to \infty)$

and $|\lambda| < 1$ for all eigenvalues of Q.

This in turn implies $I - Q$ is invertible $\left(\text{no } \lambda = 1 \Leftrightarrow \det(I-Q) \neq 0\right)$

Let $M = (I-Q)^{-1}$. This matrix will provide info about the # of steps expected to take to go from a transient state to a recurrent class, as well as the # of visits to each transient state before getting absorbed.

First we need a linear algebra result relating sums of powers of a matrix to matrices of form $(I-A)^{-1}$.

<u>Lemma 3.10</u> $\sum_{n=0}^{\infty} A^n = (I-A)^{-1}$ if $A^n \to 0$ as $n \to \infty$

Proof: Check $(I-A)(I+A+A^2+\cdots) = I$

Now let's see what $M = (I-Q)^{-1}$ can tell us about visits to transient states.

Let i be a transient state.

Let Y_i = total # visits to i = $\sum_{n=0}^{\infty} \underbrace{I\{X_n = i\}}_{\text{indicator fn} = \begin{cases} 1 & \text{if } X_n = i \\ 0 & \text{if } X_n \neq i \end{cases}}$

$\mathbb{E}[I\{X_n=i\}] = \mathbb{P}\{X_n = i\}$

Start in some transient state j, $X_0 = j$

$$\mathbb{E}[Y_i \mid X_0 = j] = \mathbb{E}\left[\sum_{n=0}^{\infty} I\{X_n = i\} \mid X_0 = j\right]$$

$$= \sum_{n=0}^{\infty} \mathbb{P}\{X_n = i \mid X_0 = j\}$$

$$= \sum_{n=0}^{\infty} \underbrace{P^n_{ji}}_{\text{both transient states in } T} = \sum_{n=0}^{\infty} Q^n_{ji} = (I-Q)^{-1}_{ji} = M_{ji}$$

$\Rightarrow M_{ji}$ = expected # visits to transient state i if start at transient state j

M_{ji} = expected # visits to i if start at j (i & j transient states) before MC is absorbed into a recurrent class

$\sum_{i \in T} M_{ji}$ = expected # steps before MC is absorbed into a recurrent class if start at transient state j

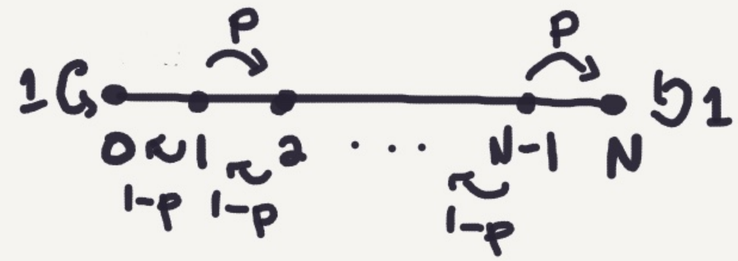
(summed over transient states)

$\boxed{\text{R demo}}$ 7-state example to practice

Common strategy: make a particular state of interest absorbing so all other states are transient (to obtain Q)

- # Analysis of Random Walk on a Circle

First we need a Gambler's Ruin result



Let a_j = prob of getting absorbed into state N if start at state j

$a_0 = 0$ and $a_N = 1$.

For $0 < j < N$, $a_j = (1-p)\,a_{j-1} + p\,a_{j+1}$

(take one step forward)

If $p = 1/2$, then $\boxed{a_j = j/N}$

If $p \neq 1/2$, then seek sol'n of form x^j: $x^j = (1-p)\,x^{j-1} + p\,x^{j+1}$

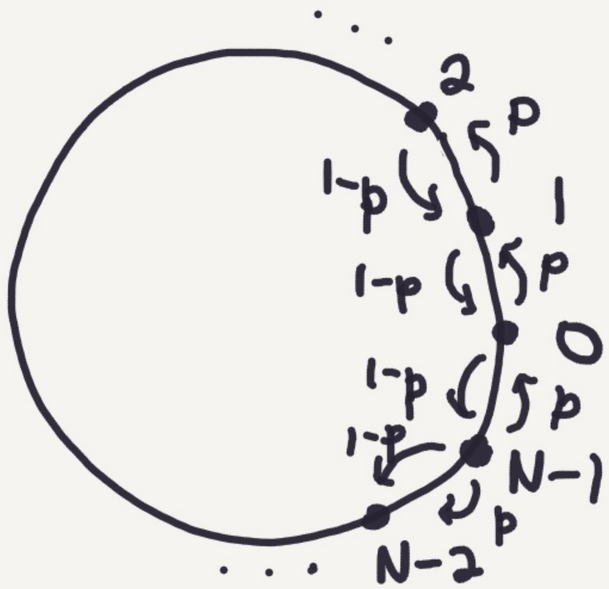
$$\Rightarrow p x^2 - x + (1-p) = 0$$

$$\Rightarrow x = \frac{1 \pm \sqrt{1-4p(1-p)}}{2p} = \frac{1 \pm (1-2p)}{2p} = 1,\ \frac{1-p}{p}$$

So $a_j = c_1 + c_2\left(\frac{1-p}{p}\right)^j$, $a_0 = 0$, $a_N = 1$,

leading to $\boxed{a_j = \dfrac{1-\left(\frac{1-p}{p}\right)^j}{1-\left(\frac{1-p}{p}\right)^N}}$

Apply to Random Walk on a Circle



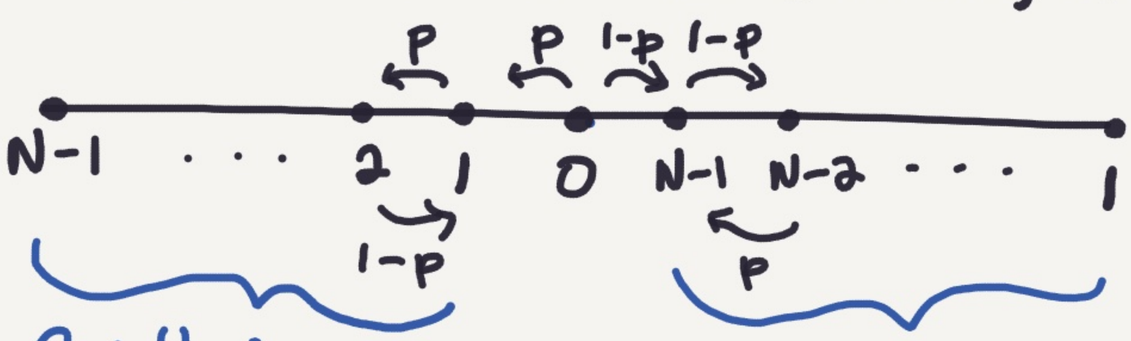
Assume $N \geq 2$ and $X_0 = 0$

$p = \mathbb{P}\{\text{take a step CCW}\}$

$1-p = \mathbb{P}\{\text{take a step CW}\}$

We want to find the prob of visiting all other states before returning to state 0

Treat as double Gambler's Ruin, conditioning on first step:



Gambler's ruin, want to hit $N-1$ before 0, starting at 1

Gambler's ruin, want to hit 1 before 0, starting at $N-1$

(finish in Monday's lab)

Markov Chain Review

A sequence of random variables $X_0, X_1, X_2, \cdots$ is called a Markov chain if $\mathbb{P}\{X_n = j \mid X_0 = x_0, X_1 = x_1, \cdots, X_{n-1} = i\}$

$$= \mathbb{P}\{X_n = j \mid X_{n-1} = i\}$$

for all $n \geq 1$ and states i, j.

That is, the chain is governed by one-step transition probabilities, which can be represented by a transition matrix P if the state space is finite.

Powers P^n_{ij} yield the n-step transition prob $\mathbb{P}\{X_n = j \mid X_0 = i\}$

Limiting distribution $\bar{\lambda}$: $\lim_{n\to\infty} P^n_{ij} = \lambda_j$ for all states i & j

(doesn't depend on initial state i)

Equivalently, $\lim_{n\to\infty} P^n = \begin{bmatrix} \bar{\lambda} \\ \vdots \\ \bar{\lambda} \end{bmatrix}$

(all rows converge to $\bar{\lambda}$)

Stationary distribution $\bar{\pi}$: $\bar{\pi} = \bar{\pi} P$

(also called invariant distribution)

$\pi_j = \sum_i \pi_i P_{ij}$ for all states j

Limiting distributions are stationary distributions,

but not vice versa.

Summary of matrix methods for chain computations

① For irreducible, aperiodic MC, the long-term probability of visiting each state is given by π (left eigenvector for eigenvalue 1)
The first return time for state j is given by $1/\pi_j$

② For a reducible MC, organize transition matrix $P = \begin{array}{c|c} & \begin{array}{cc} t & r \end{array} \\ \hline \begin{array}{c} t \\ r \end{array} & \left[\begin{array}{c:c} Q & R \\ \hdashline 0 & \tilde{P} \end{array}\right] \end{array}$

For transient states, $M = (I-Q)^{-1}$ has entries

M_{ji} = expected # of visits to transient state i if start at transient state j

Sum j^{th} row of M to obtain total expected # of visits to transient states if start at transient state j

③ To apply to irreducible chain, we can compute the expected # of visits to state i before hitting state k if start at j by turning k into an absorbing state (so rest of states become transient). Then use matrix M.

④ For irreducible MC, to find the probability of hitting state b before state c, let b & c be absorbing states.

$$P = \begin{array}{c} \\ t \\ r \end{array} \begin{array}{c} \begin{array}{cc} & \begin{array}{cc} b & c \end{array} \end{array} \\ \left[\begin{array}{c|c} Q & R \\ \hline 0 & I \end{array}\right] \end{array}$$

Let A be the matrix with entries $A_{ij} = \mathbb{P}\{X_n = j \text{ for some } n \geq 1 : X_0 = i\}$

For transient state i and absorbing state j,

$$A_{ij} = \mathbb{P}\{X_n = j \text{ eventually} \mid X_0 = i\}$$

$$= \sum_{\text{all states } k} \mathbb{P}\{X_1 = k \mid X_0 = i\} \, \mathbb{P}\{X_n = j \text{ eventually} \mid X_1 = k\}$$

$$= \sum_{\text{transient } k} \underbrace{\mathbb{P}\{X_1 = k \mid X_0 = i\}}_{Q_{ik}} \underbrace{\mathbb{P}\{X_n = j \text{ eventually} \mid X_1 = k\}}_{A_{kj}}$$

$$+ \sum_{\text{recurrent } k} \underbrace{\mathbb{P}\{X_1 = k \mid X_0 = i\}}_{R_{ik}} \underbrace{\mathbb{P}\{X_n = j \text{ eventually} \mid X_1 = k\}}_{= \begin{cases} 1 & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases}}$$

$$= \sum_{\text{transient } k} Q_{ik} A_{kj} + R_{ij} \Rightarrow A = QA + R$$

$$\Rightarrow (I - Q)A = R$$

$$\Rightarrow A = (I - Q)^{-1} R = MR$$

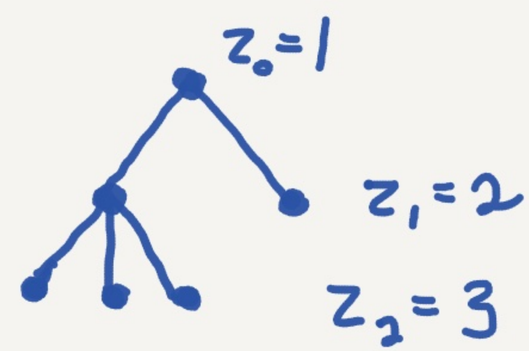
Chapter 4 Branching Processes

- stochastic model for population growth
- late 1800s, Galton & Watson developed to track male surnames from generation to generation in England.
- historical note: Galton was an influential statistician who created the concept of correlation. He also advocated social Darwinism, eugenics, and scientific racism.
  Watson was a mathematician working on electricity & magnetism, who was also an ordained priest in England.

Def'n: A <u>branching process</u> is a sequence of RVs $Z_0, Z_1, Z_2, \cdots$ with a state space $\{0,1,2,\cdots\}$, where Z_n = # individuals in generation n, in which each individual independently produces offspring according to an offspring distribution $a = (a_0, a_1, a_2, \cdots)$, $0 \le a_k \le 1$, $\sum_{k=0}^{\infty} a_k = 1$,

a_k = prob of k offspring

Typically start with $Z_0 = 1$:



Note that 0 will be an absorbing state (extinction)

What can we say about the population size over time?

If $a_0 = 0$, then always positive. If $a_0 = 1$, instant extinction.

Will usually assume $0 < a_0 < 1$ <u>and</u> $a_0 + a_1 < 1$, so positive prob of multiple offspring.

Lemma 4.1 If $0 < a_0 < 1$ and $a_0 + a_1 < 1$, then all nonzero states are transient.

Proof: Note that prob of going from i to 0 is $(a_0)^i$ all i individuals had 0 offspring

Calculate f_i = prob of eventually hitting i again if start at i

($f_i < 1$ implies i transient)

$$
\begin{aligned}
f_i &= \mathbb{P}\{Z_n = i \text{ for some } n \geq 1 \mid Z_0 = i\} \\
&\leq \mathbb{P}\{Z_1 > 0 \mid Z_0 = i\} \\
&= 1 - \mathbb{P}\{Z_1 = 0 \mid Z_0 = i\} \\
&= 1 - (a_0)^i \\
&< 1
\end{aligned}
$$

$\Rightarrow$ state i is transient for $i \geq 1$.

All positive states being transient in an infinite state space implies the population either goes extinct or explodes.

Section 4.2 Mean generation size

Let X_i = # offspring of ith individual in $(n-1)^{st}$ generation (iid with distribution a)

Note $Z_n = \sum_{i=1}^{Z_{n-1}} X_i$ (sum offspring over the individuals in $(n-1)^{st}$ generation)

Mean of offspring distribution: $\mu = \sum_{k=0}^{\infty} k a_k = \mathbb{E}[X_i]$

Variance is $\sigma^2 = \sum_{k=0}^{\infty} a_k (k-\mu)^2$

Expected value of nth generation is then

$$\mathbb{E}[Z_n] = \sum_{k=0}^{\infty} \mathbb{E}[Z_n \mid Z_{n-1} = k]\, \mathbb{P}\{Z_{n-1} = k\}$$

$$= \sum_{k=0}^{\infty} \mathbb{E}\left[\sum_{i=1}^{k} X_i\right] \mathbb{P}\{Z_{n-1} = k\}$$

$$= \sum_{k=0}^{\infty} k\mu\, \mathbb{P}\{Z_{n-1} = k\} = \mu \sum_{k=0}^{\infty} k\, \mathbb{P}\{Z_{n-1} = k\}$$

$$= \mu\, \mathbb{E}[Z_{n-1}]$$

If $Z_0 = 1$, then $\mathbb{E}[Z_n] = \mu^n$.

So if $Z_0 = 1$, then $\mathbb{E}[Z_n] = \mu^n$.

(In general, $\mathbb{E}[Z_n] = \mu^n \mathbb{E}[Z_0]$.)

Three cases can occur:

1. Subcritical $\mu < 1$

$$\mathbb{P}\{Z_n = 0\} = 1 - \mathbb{P}\{Z_n \geq 1\} \geq 1 - \mathbb{E}[Z_n] = 1 - \mu^n \to 1 \text{ as } n \to \infty$$

2. Critical $\mu = 1$

3. Supercritical $\mu > 1$

$\boxed{\text{R demo}}$

When can extinction occur? When can population explosions occur?

Variance of generation size, assuming $Z_0 = 1$

$$\mathrm{Var}[Z_n] = \mathrm{Var}\left[\mathbb{E}[Z_n \mid Z_{n-1}]\right] + \mathbb{E}\left[\mathrm{Var}[Z_n \mid Z_{n-1}]\right]$$ Law of Total Variance

$$= \mathrm{Var}[\underbrace{\mu}_{\mathbb{E}[a]} Z_{n-1}] + \mathbb{E}[\underbrace{\sigma^2}_{\mathrm{Var}[a]} Z_{n-1}]$$ using independence of X_i

$$= \mu^2 \mathrm{Var}[Z_{n-1}] + \sigma^2 \mu^{n-1}$$

If $Z_0 = 1$, then $\mathrm{Var}[Z_0] = 0$

$$\mathrm{Var}[Z_1] = \sigma^2$$

$$\mathrm{Var}[Z_2] = \sigma^2\mu^2 + \sigma^2\mu = \sigma^2\mu(1+\mu)$$

$$\vdots$$

$$\mathrm{Var}[Z_n] = \sigma^2 \mu^{n-1} \sum_{k=0}^{n-1} \mu^k = \begin{cases} n\sigma^2 & \text{if } \mu = 1 \\ \sigma^2 \mu^{n-1}\left(\frac{\mu^n - 1}{\mu - 1}\right) & \text{if } \mu \neq 1 \end{cases}$$

(Proof by induction)

So if $Z_0 = 1$, then $\mathbb{E}[Z_n] = \mu^n$ and

$$\mathrm{Var}[Z_n] = \sigma^2 \mu^{n-1} \sum_{k=0}^{n-1} \mu^k = \begin{cases} n\sigma^2 & \text{if } \mu = 1 \\ \sigma^2 \mu^{n-1} \left(\frac{\mu^n - 1}{\mu - 1} \right) & \text{if } \mu \neq 1 \end{cases}$$

① Subcritical $\mu < 1$: $\lim_{n\to\infty} \mathbb{E}[Z_n] = 0$

$\lim_{n\to\infty} \mathbb{P}\{Z_n = 0\} = 1$

$\lim_{n\to\infty} \mathrm{Var}[Z_n] = 0$

② Critical $\mu = 1$: $\lim_{n\to\infty} \mathbb{E}[Z_n] = 1$ but $\mathrm{Var}[Z_n] = n\sigma^2 \to \infty$ as $n \to \infty$

(enormous variability in pop size)

③ Supercritical $\mu > 1$: $\lim_{n\to\infty} \mathbb{E}[Z_n] = \infty$

$\lim_{n\to\infty} \mathrm{Var}[Z_n] = \infty$

(growing exponentially fast as $n \to \infty$)

Can get either extinction or exponential pop growth in case ③.
What are the probabilities of each, depending on offspring distribution a?
(Cases ① & ② always go extinct)

Expected value of # of individuals in n^{th} generation is

$$\mathbb{E}[Z_n] = \mu^n \mathbb{E}[Z_0].$$

But the variance can be enormous:

① Subcritical $\mu < 1$: $\lim_{n\to\infty} \mathbb{E}[Z_n] = 0$

$\lim_{n\to\infty} \mathbb{P}\{Z_n = 0\} = 1$

$\lim_{n\to\infty} \mathrm{Var}[Z_n] = 0$

② Critical $\mu = 1$: $\lim_{n\to\infty} \mathbb{E}[Z_n] = 1$ but $\mathrm{Var}[Z_n] = n\sigma^2 \to \infty$ as $n \to \infty$

③ Supercritical $\mu > 1$: $\lim_{n\to\infty} \mathbb{E}[Z_n] = \infty$

$\lim_{n\to\infty} \mathrm{Var}[Z_n] = \infty$

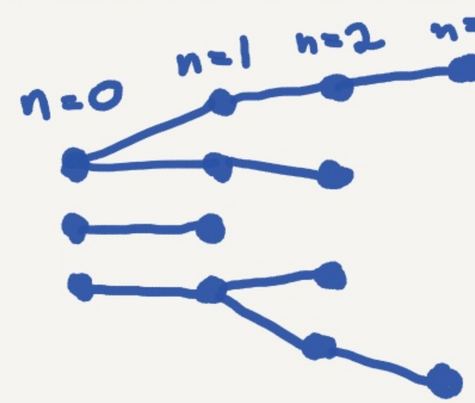
Section 4.3 Probability generating functions

Consider a branching process with $0 < a_0 < 1$ and $a_0 + a_1 < 1$

(so positive probability of ≥ 2 offspring as well as 0 offspring)

Let $s_n(k) = \mathbb{P}\{Z_n = 0 \mid Z_0 = k\}$ = prob of extinction by n^{th} generation if start with k individuals

Let $s(k) = \lim_{n \to \infty} s_n(k)$ = prob of extinction if start with k individuals



Branches act independently, so $s(k) = (s(1))^k$

We only need to solve for $s(1)$ to obtain $s(k)$.

Let $s = s(1)$ = prob of eventual extinction if start with k individuals.

$$s = \mathbb{P}\{Z_n = 0 \text{ for some } n \geq 1 \mid Z_0 = 1\}$$

$$= \sum_{k=0}^{\infty} \mathbb{P}\{\text{eventual extinction} \mid Z_1 = k\}\, \mathbb{P}\{Z_1 = k \mid Z_0 = 1\}$$

$$= \sum_{k=0}^{\infty} s(k) a_k$$

$$= \sum_{k=0}^{\infty} a_k s^k,$$ which is the probability generating function for the offspring distribution $(a_0, a_1, a_2, \cdots)$

$$= G(s)$$

The solution to $s = G(s)$ gives the prob s of eventual extinction.

We'll call this solution e.

Background on probability generating functions

Defn: For a discrete RV X taking values in $\{0,1,2,\cdots\}$, the <u>prob generating function</u> of X is defined to be

$$G(s) = \mathbb{E}[s^X] = \sum_{k=0}^{\infty} s^k \, \mathbb{P}\{X=k\},$$

which converges absolutely for $|s| \le 1$ (since $\sum_{k=0}^{\infty} \mathbb{P}\{X=k\} = 1$)

Note: $G(1) = \mathbb{E}[1^X] = \sum_{k=0}^{\infty} 1 \, \mathbb{P}\{X=k\} = 1$

$G(0) = \mathbb{P}\{X=0\}$

$G'(s) = \sum_{k=1}^{\infty} k s^{k-1} \mathbb{P}\{X=k\} = \mathbb{P}\{X=1\} + 2s\,\mathbb{P}\{X=2\} + 3s^2\,\mathbb{P}\{X=3\} + \cdots$

so $G'(s) \ge 0$ for $0 \le s \le 1$ ($G(s)$ is an increasing function of s)

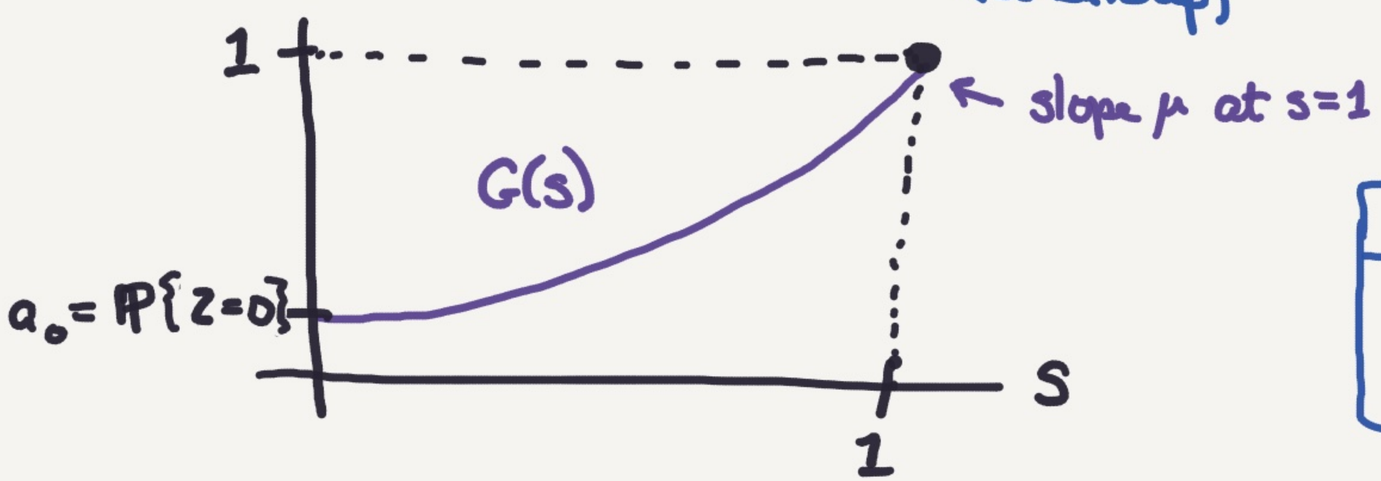
$G'(1) = \sum_{k=1}^{\infty} k\,\mathbb{P}\{X=k\} = \sum_{k=0}^{\infty} k\,\mathbb{P}\{X=k\} = \mathbb{E}[X] = \mu$

$G''(s) = \sum_{k=2}^{\infty} k(k-1) s^{k-2} \mathbb{P}\{X=k\}$, so $G''(s) > 0$ if $0 < s \le 1$ and $\mathbb{P}\{X \ge 2\} > 0$

Summary: $G(1)=1$, $G(0)=\underbrace{\mathbb{P}\{Z=0\}}_{a_0}$, $G'(s)\geq 0$ for $0\leq s\leq 1$,

$G'(1)=\mu$, $G''(s)>0$ for $0<s\leq 1$ if $\underbrace{\mathbb{P}\{Z\geq 2\}>0}_{a_0+a_1<1}$.

(concave up)



$$\boxed{\text{The solution to } s=G(s) \text{ gives the extinction prob } e.}$$

Example: $a_0=1/4$ $a_1=1/2$ $a_2=1/4$ $a_n=0$ for $n>2$

$$\mu = 0\cdot\tfrac{1}{4}+1\cdot\tfrac{1}{2}+2\cdot\tfrac{1}{4}=1$$

$$G(s)=\frac{1}{4}+\frac{1}{2}s+\frac{1}{4}s^2$$

The root of $s=G(s)$ with $0\leq s\leq 1$ gives the extinction probability:

$$s^2-2s+1=0$$
$$(s-1)^2=0 \Rightarrow e=1$$

Always goes eventually extinct

Example: $a_0 = 1/4 \quad a_1 = 1/4 \quad a_2 = 1/2 \quad a_n = 0$ for $n > 2$

$$\mu = 0 \cdot \tfrac{1}{4} + 1 \cdot \tfrac{1}{4} + 2 \cdot \tfrac{1}{2} = \tfrac{5}{4}$$

$$G(s) = \tfrac{1}{4} + \tfrac{1}{4} s + \tfrac{1}{2} s^2$$

$\boxed{\text{R demo}}$

Solve $s = G(s)$: $2s^2 - 3s + 1 = 0 \Rightarrow s = 1, 1/2$

Which root? We know the population has a chance of exploding population, so $e = 1/2$. (Will have a better way of deciding in a bit.)

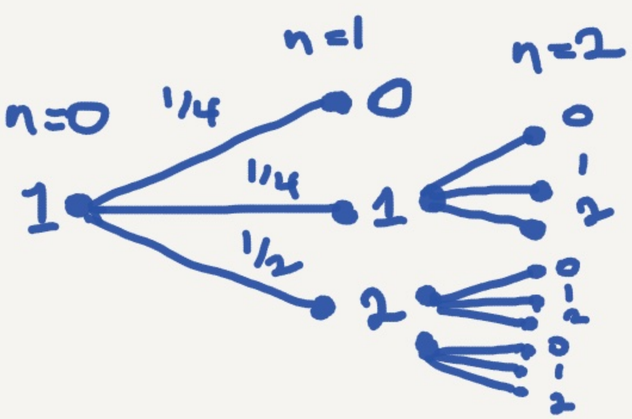
Let $Z_0 = 1$.

What is probability of going extinct at generation $n = 1$?

What is prob of going extinct at $n = 2$?

Direct calculation: $\mathbb{P}\{Z_2 = 0 \mid Z_1 > 0\} = \dfrac{\mathbb{P}\{Z_2 = 0 \text{ and } Z_1 > 0\}}{\mathbb{P}\{Z_1 > 0\}}$ $(Z_0 = 1)$

$$= \frac{1/4 \cdot 1/4 + 1/2 (1/4)^2}{1/4 + 1/2} = 1/8$$



Alternative: $\mathbb{P}\{Z_2 = 0 \mid Z_1 > 0\}$

$$= \frac{\mathbb{P}\{Z_2 = 0\} - \mathbb{P}\{Z_1 = 0\}}{1 - \mathbb{P}\{Z_1 = 0\}} = \frac{s_2(1) - s_1(1)}{1 - s_1(1)}$$

(assuming $Z_0 = 1$)

We want a means of calculating $s_1(1)$ and $s_2(1)$,

as well as knowing which root of $s = G(s)$ is the extinction prob.

Assume $Z_0 = 1$.

Let $G_0(s) = s$ (generating fn for Z_0)

$$G_1(s) = G(s) = \sum_{k=0}^{\infty} a_k s^k \qquad \text{(generating fn for } Z_1\text{)}$$

$$G_2(s) = \sum_{k=0}^{\infty} s^k \mathbb{P}\{Z_2 = k\} \qquad \text{(generating fn for } Z_2\text{)}$$

$$G_n(s) = \sum_{k=0}^{\infty} s^k \mathbb{P}\{Z_n = k\}$$

Note: $G_n(0) = \mathbb{P}\{Z_n = 0 \mid Z_0 = 1\} = s_n(1)$

Iterative Formula: $G_n(s) = G_1(G_{n-1}(s))$, so $s_1(1) = a_0 = G_1(0)$

$s_2(1) = G_2(0) = G_1(G_1(0)) = G(a_0)$

Iterative Formula: $G_n(s) = G_1(G_{n-1}(s))$

Proof: $G_n(s) = \sum_{k=0}^{\infty} s^k \mathbb{P}\{Z_n = k \mid Z_0 = 1\}$

$$= \sum_{k=0}^{\infty} s^k \left(\sum_{j=0}^{\infty} \mathbb{P}\{Z_n = k \mid Z_1 = j\} \mathbb{P}\{Z_1 = j \mid Z_0 = 1\} \right)$$

$$= \sum_{j=0}^{\infty} \mathbb{P}\{Z_1 = j \mid Z_0 = 1\} \sum_{k=0}^{\infty} s^k \mathbb{P}\{Z_n = k \mid Z_1 = j\}$$

$$= \sum_{j=0}^{\infty} a_j \sum_{k=0}^{\infty} s^k \mathbb{P}\{Z_{n-1} = k \mid Z_0 = j\} = \sum_{j=0}^{\infty} a_j G_{n-1}(s)^j$$

Note that each branch independently generates offspring

$$G_Y(s) = \mathbb{E}[s^Y] = \sum_{k=0}^{\infty} \mathbb{P}\{Y = k\} s^k$$

$$G_{Y_1 + \cdots + Y_j}(s) = \mathbb{E}[s^{Y_1 + \cdots + Y_j}] = \mathbb{E}[s^{Y_1} s^{Y_2} \cdots s^{Y_j}]$$

by independence

$$= \mathbb{E}[s^{Y_1}] \mathbb{E}[s^{Y_2}] \cdots \mathbb{E}[s^{Y_j}]$$

$$= G_{Y_1}(s) G_{Y_2}(s) \cdots G_{Y_j}(s)$$

$$= (G_Y(s))^j$$

So $G_n(s) = G(G_{n-1}(s))$.

Return to example:

$a_0 = 1/4 \quad a_1 = 1/4 \quad a_2 = 1/2$

$$G(s) = \frac{1}{4} + \frac{1}{4}s + \frac{1}{2}s^2$$

$$s_1(1) = a_0 = 1/4$$

$$s_2(1) = G(a_0) = G(1/4) = \frac{1}{4} + \frac{1}{4}\cdot\frac{1}{4} + \frac{1}{2}\left(\frac{1}{4}\right)^2 = \frac{11}{32}$$

$$\mathbb{P}\{Z_2 = 0 \mid Z_1 > 0\} = \frac{s_2(1) - s_1(1)}{1 - s_1(1)}$$

$$= \frac{11/32 - 1/4}{1 - 1/4} = 1/8$$

Example: $a_0 = 0.5 \quad a_1 = 0.1 \quad a_2 = 0.4 \quad a_n = 0$ for $n \geq 3$

$$\mu = 0(0.5) + 1(0.1) + 2(0.4) = 0.9 \Rightarrow e = 1$$ subcritical case

Find $\mathbb{P}\{Z_3 = 0 \mid Z_2 > 0, Z_0 = 1\}$

$$= \frac{\mathbb{P}\{Z_3 = 0 \mid Z_0 = 1\} - \mathbb{P}\{Z_2 = 0 \mid Z_0 = 1\}}{\mathbb{P}\{Z_2 = 0 \mid Z_0 = 1\}}$$

$$= \frac{s_3(1) - s_2(1)}{1 - s_2(1)}$$

$s_1(1) = a_0 = 0.5$

$s_2(1) = G(s_1(1)) = G(.5) = 0.5 + 0.1(0.5) + 0.4(0.5)^2 = 0.65$

$s_3(1) = G(s_2(1)) = G(0.65) = 0.734$

So $\mathbb{P}\{Z_3 = 0 \mid Z_2 > 0, Z_0 = 1\} = \dfrac{0.734 - 0.65}{1 - 0.65} = 0.24$

We can also show that the extinction prob

$e = \lim_{n\to\infty} s_n(1)$ is the smallest positive root of $s = G(s)$.

Let $\hat{s}$ be the smallest positive root of $s = G(s)$.

We claim $s_n(1) = \mathbb{P}\{Z_n = 0 \mid Z_0 = 1\} \leq \hat{s}$ for all $n \geq 0$.

Base case $n=0$: $s_0(1) = 0 \leq \hat{s}$.

Induction step: Assume $s_{n-1}(1) \leq \hat{s}$.

Then $s_n(1) = G_n(0) = G_1(G_{n-1}(0)) = G(s_{n-1}(1))$

$G(s)$ is increasing ($G'(s) \geq 0$ for $0 \leq s \leq 1$),

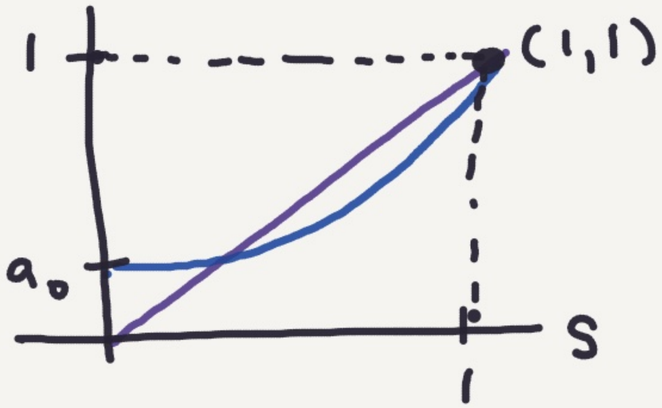
so $s_n(1) = G(s_{n-1}(1)) \leq G(\hat{s}) = \hat{s}$ (recall $\hat{s}$ is root of $s = G(s)$)

Hence $s_n(1) \leq \hat{s}$ for all $n \geq 0$,

implying that $e = \lim_{n\to\infty} s_n(1) \leq \hat{s}$.

The extinction prob e is a positive root of $s = G(s)$ and can't be smaller than the smallest positive root $\hat{s}$, so we must have $e = \hat{s}$.

To visualize what is happening in our 3 cases,



recall that $G(0)=a_0$, $G(1)=1$,
$G'(s)\geq 0$ for $0\leq s\leq 1$, $G'(1)=\mu$,
and $G''(s)>0$ for $0<s\leq 1$ if $a_0+a_1<1$

$\boxed{\text{Mathematica demo}}$

Also note that $e=\lim_{n\to\infty} s_n(1)$ is a fixed point of the iteration $s_n=G(s_{n-1})$, which generates a sequence converging to e.

Generally, $\mathbb{P}\{Z_n=0 \text{ and } Z_{n-1}>0\} = s_n(1)-s_{n-1}(1)$
$= G_n(0)-G_{n-1}(0)$

prob of going extinct at the nth generation (and not before)

Section 4.4 Extinction is forever

Recall $G(s) = \sum_{k=0}^{\infty} a_k s^k$ generating fn for Z_1

$$G_n(s) = \sum_{k=0}^{\infty} \mathbb{P}\{Z_n = k\} s^k$$ generating fn for Z_n

$$= G(G_{n-1}(s))$$

<u>Thm 4.2</u> For a branching process with prob gen fn $G(s)$, the probability of eventual extinction is the smallest positive root of $s = G(s)$.

If $\mu \leq 1$ (subcritical or critical), then $e = 1$.

What about the expected time to extinction?

Given a branching process with offspring distribution

$a = (a_0, a_1, a_2, \dots)$ and $\mu \leq 1$ (so prob 1 of going extinct),

the expected value of extinction time T (given $Z_0 = 1$) is

$$\mathbb{E}[T] = \sum_{n=1}^{\infty} n \, \mathbb{P}\{T = n \mid Z_0 = 1\}$$

$$= \sum_{n=1}^{\infty} \sum_{k=1}^{n} \mathbb{P}\{T = n \mid Z_0 = 1\}$$

$$= \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} \mathbb{P}\{T = n \mid Z_0 = 1\}$$

$$= \sum_{k=1}^{\infty} \mathbb{P}\{\underbrace{T \geq k}_{Z_{k-1} > 0} \mid Z_0 = 1\}$$

(new index $n = k-1$)

$$= \sum_{n=0}^{\infty} \mathbb{P}\{Z_n > 0 \mid Z_0 = 1\} = 1 + \sum_{n=1}^{\infty} \mathbb{P}\{Z_n > 0 \mid Z_0 = 1\}$$

$$= 1 + \sum_{n=1}^{\infty} (1 - s_n(1))$$

When does this series converge?

Ratio test:

$$\lim_{n \to \infty} \frac{1 - s_{n+1}(1)}{1 - s_n(1)} = \lim_{n \to \infty} \frac{G(1) - G(s_n(1))}{1 - s_n(1)}$$

$$= G'(1) = \mu$$

$\Rightarrow$ converges to finite value for $\mu < 1$

$$\mathbb{E}[T] = 1 + \sum_{k=1}^{\infty} (1 - s_n(1)) \text{ for } \mu \leq 1$$

When does this series converge?

$$\boxed{\begin{aligned} G_n(s) &= \sum_{k=0}^{\infty} s^k \mathbb{P}\{Z_n = k \mid Z_0 = 1\} \\ s_n(j) &= \mathbb{P}\{Z_n = 0 \mid Z_0 = j\} \end{aligned}}$$

Recall $s_1(1) = a_0$, $G(1) = 1$,

$$s_n(1) = G_n(0) = G(G_{n-1}(0)) = G(s_{n-1}(1)) \text{ for } n \geq 2$$

Ratio test: $\lim_{n\to\infty} \frac{1 - s_{n+1}(1)}{1 - s_n(1)} = \lim_{n\to\infty} \frac{G(1) - G(s_n(1))}{1 - s_n(1)} = G'(1) = \mu$

So the series converges for $\mu < 1$, that is $\mathbb{E}[T] < \infty$ for $\mu < 1$ (subcritical)

It can be shown that $\mathbb{E}[T] = \infty$ for the $\mu = 1$ case (critical)

(will see example in branching processes lab)

(For the supercritical case, $\mu > 1$, we have $e < 1$, so can't compute $\mathbb{E}[T]$.)

Chapter 5 Markov Chain Monte Carlo (MCMC)

- powerful sampling algorithms, associated with Bayesian statistics

Given a probability distribution π, the goal of MCMC methods is to simulate a random variable X whose distribution is π.

The idea is to construct an appropriate ergodic MC with limiting distr. π. Run that Markov chain long enough to converge to that long-term π ("burn-in"), then start using the samples generated by the chain.

Recall Law of Large Numbers: given iid sequence $Y_1, Y_2, Y_3, \cdots$ of RVs with common mean $\mu < \infty$, we have $\lim_{n\to\infty} \frac{1}{n}(Y_1 + \cdots + Y_n) = \mu$ with prob 1.

(That is, the sample mean approaches the true mean as sample size $\to \infty$.)

Also works for bounded functions: $\lim_{n\to\infty} \frac{f(Y_1) + \cdots + f(Y_n)}{n} = \mathbb{E}[f(Y)]$

Theorem 5.1 Strong law of large numbers for Markov chains

Suppose $X_0, X_1, \cdots$ is an $\underbrace{\text{ergodic}}_{\text{irreducible, aperiodic, all states have finite expected return times}}$ MC with stationary distr π and state space S.

Let f be a bounded, real-valued function and X an RV with distr π.

Then with prob 1, $$\lim_{n\to\infty} \frac{f(X_0)+\cdots+f(X_{n-1})}{n} = \mathbb{E}[f(X)] = \sum_{j\in S} f(j)\pi_j$$

Note: RV X is distinct from RVs in chain $X_0, X_1, \cdots$, but shares same π.

Markov chains are <u>not</u> independent sequences (X_{n+1} depends on X_n), but it is true that, for ergodic chains, successive excursions between visits to the same state are independent (like restarting chain).

Given an ergodic Markov chain $X_0, X_1, \cdots$ with stationary distribution π, let A be a nonempty set of states and set $\pi_A = \sum_{j \in A} \pi_j$ (will show this is the long-term expected proportion of visits to set A)

Let $I_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$ (indicator fn for set A)

Then $\sum_{k=0}^{n-1} I_A(X_k) = \#$ visits to states in set A in steps 0 to $n-1$.

According to Thm 5.1, with prob 1, $\lim_{n \to \infty} \frac{I_A(X_0) + \cdots + I_A(X_{n-1})}{n}$

$$= \mathbb{E}[I_A(X)] = \mathbb{P}\{X \in A\}$$

$$= \pi_A$$

So the long-term expected proportion of visits to states in set A is π_A.

Example: binary sequence with no adjacent 1s

00101000 10100010 ...

We want an algorithm that samples a sequence of length m of 0s and 1s with no adjacent 1s (call this a "good" sequence).

Challenge: "good" sequences are rare in the set of binary sequences.

$m=4$: $2^4 = 16$ possible binary sequences, with 8 "good" seq.

$m=100$: $2^{100} \approx 10^{30}$ possible binary seq, with 10^{21} "good" seq

(prob 10^{-9} of randomly choosing a good seq)

Suppose you want to efficiently generate a randomly chosen good seq.

1) Brute force: store all possible good sequences, randomly choose one.

2) Rejection method: generate random binary sequences through m coin flips, reject if not good seq, repeat until obtain good seq.

3) MCMC approach: construct ergodic Markov chain $X_0, X_1, \ldots$ whose state space is the set of good sequences and whose limiting distribution is uniform.

don't need to store, unlike brute force approach

MCMC approach: construct ergodic Markov chain $X_0, X_1, \ldots$ whose state space is the set of good sequences and whose limiting distribution is uniform.

Algorithm: Start with a good sequence as X_0 state (of length m)

Pick one of m components at random

If component is 1, flip to 0 (still a good seq)

If component is 0, switch to 1 if results in a good seq, otherwise leave as 0.

$\boxed{\text{R demo}}$

The underlying idea here is that of the Metropolis-Hastings algorithm.

Section 5.2 Metropolis-Hastings Algorithm

developed in 1950s

extended in 1970

Given a discrete prob distribution $\bar{\pi}$ ("target distr"), we want to construct a reversible Markov chain $X_0, X_1, \cdots$ whose stationary distribution is $\bar{\pi}$.

(Reversible if $\pi_i P_{ij} = \pi_j P_{ji}$ for all states i & j.
Time-reversal chain is $\tilde{P}_{ij} = \dfrac{\pi_j P_{ji}}{\pi_i}$.)

Let T be any transition matrix on the same state space as $X_0, X_1, \cdots$, Choosing T to represent a Markov chain that is easy to work with. It will be used to generate "proposal states" (like flipping a random bit in the binary seq example).

Algorithm then decides whether or not to accept the proposal state (or stay at the current state).

Metropolis-Hastings Algorithm

Suppose we're at the nth step of the chain with $X_n = i$.

① Choose a $\underbrace{\text{state } j}_{\text{proposal state}}$ from the state space with probability $\underbrace{T_{ij}}_{\text{proposal matrix}}$

② Decide whether to accept the proposal state:

Let $a(i,j) = \dfrac{\pi_j T_{ji}}{\pi_i T_{ij}}$ acceptance function

If $a(i,j) \geq 1$, accept the proposal state j

Otherwise accept j with probability $a(i,j)$ (keep i if reject j)

That is, $X_{n+1} = \begin{cases} j & \text{if } U \leq a(i,j) \\ i & \text{if } U > a(i,j) \end{cases}$ where $U \sim \text{Unif}(0,1)$

<u>Theorem</u> The sequence $X_0, X_1, \cdots$ generated by the MH algorithm is a reversible Markov chain with stationary distribution π.

Proof: Markov because X_{n+1} only depends on X_n.

Let P be the transition matrix resulting from the MH algorithm applied to the proposal matrix T.

We need to prove $\pi_i P_{ij} = \pi_j P_{ji}$ for all states i & j, $i \neq j$.

Prob of j being the proposal state is T_{ij}

Prob of j being accepted is $\min(a(i,j), 1)$

$$\mathbb{P}\{U \leq a(i,j)\} = \begin{cases} a(i,j) & \text{if } a(i,j) \leq 1 \\ 1 & \text{if } a(i,j) > 1 \end{cases}$$

uniform RV on $[0,1]$

$$\text{So } P_{ij} = \begin{cases} T_{ij}\, a(i,j) & \text{if } \pi_j T_{ji} \leq \pi_i T_{ij} \\ T_{ij} & \text{if } \pi_j T_{ji} > \pi_i T_{ij} \end{cases}$$

Case ①: If $\pi_j T_{ji} \leq \pi_i T_{ij}$. then $\pi_i P_{ij} = \cancel{\pi_i}\, \cancel{T_{ij}} \cdot \dfrac{\pi_j T_{ji}}{\cancel{\pi_i}\, \cancel{T_{ij}}} = \pi_j \overbrace{T_{ji}}^{=P_{ji}}$

Case ②: If $\pi_j T_{ji} > \pi_i T_{ij}$, then $\pi_i P_{ij} = \pi_i T_{ij} \cdot \dfrac{\pi_j T_{ji}}{\pi_j T_{ji}} = \pi_j T_{ji}\, a(j,i)$

$= \pi_j P_{ji}$ because $a(j,i) \leq 1$

So MH chain is reversible, with stationary distr. π.

Numerical issues with MCMC

① Burn-in: need to run chains long enough that prob distr. of visiting states has converged to target distr. π (discard first part, only keep later samples)

② Thinning: successive steps will be correlated (X_{n+1} depends on X_n) so run many chains, discard burn-in portion, then keep every k^{th} value (k depends on how long correlation takes to decay).

③ Mixing: check that not too many proposal states are rejected (could indicate poor choice of proposal transition matrix T)

MCMC software typically provides diagnostics to check these and other issues, to ensure good results.

Why MCMC works well for Bayesian methods

Let D be the observed data and θ represent model parameters

Posterior distribution

$$\underbrace{\mathbb{P}\{\theta \mid D\}}_{\text{target distr } \pi} = \frac{\overbrace{\mathbb{P}\{\theta\}}^{\text{prior}}\ \overbrace{\mathbb{P}\{D \mid \theta\}}^{\text{likelihood of the observed data being generated by the model}}}{\underbrace{\mathbb{P}\{D\}}_{\text{prob of these data being observed - usually unknown}}}$$

MH comes to the rescue because only the ratio $\frac{\pi_j}{\pi_i}$ is needed, canceling out the unknown $\mathbb{P}\{D\}$.

$\boxed{\text{R demo}}$ - decryption lab

Section 5.3 Gibbs sampler

Aimed at high-dimensional state spaces

Idea is to change one dimension at a time, conditioned on other dim.

Suppose the problem involves m variables $x_1, \cdots, x_m$

The target distribution in an m-dimensional joint density $\pi(\bar{x}) = \pi(x_1, \cdots, x_m)$

Bivariate example

Univariate normal density $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$ mean μ, variance σ^2

Let Z_1 and Z_2 be independent standard normal RVs ($\mu = 0$, $\sigma = 1$)

Let $X_1 = Z_1 + \mu_1$

$X_2 = \rho Z_1 + \sqrt{1-\rho^2}\, Z_2 + \mu_2$

That is, $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} + \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$

Covariance $\Sigma = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} 1 & \rho \\ 0 & \sqrt{1-\rho^2} \end{bmatrix} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$

Joint density $f(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left((x_1-\mu_1)^2 - 2\rho(x_1-\mu_1)(x_2-\mu_2) + (x_2-\mu_2)^2\right)}$

Conditional distr. $f(x_1 | x_2) = \frac{f(x_1, x_2)}{\int_{-\infty}^{\infty} f(y, x_2)\, dy} = \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} e^{-\frac{1}{2}\frac{(x_1-\mu_1-\rho(x_2-\mu_2))^2}{1-\rho^2}}$

→ univariate normal RV with mean $\mu_1 + \rho(x_2 - \mu_2)$ & variance $1-\rho^2$

$\boxed{\text{R demo}}$

Chapter 6 Poisson Process

- continuous time $[0,\infty)$, discrete state space $\{0,1,2,\cdots\}$
- aim is to count arrivals over time
  (customers entering a store, cars passing through EZ lane for toll)

Def'n: A <u>Poisson process with parameter λ</u> is a counting process $(N_t)_{t\geq 0}$ such that

(1) $N_0 = 0$

(2) $\forall\, t>0$, N_t follows a Poisson distribution with parameter λt

(3) $\forall\, s,t>0$, $N_{t+s}-N_s$ has same distribution as N_t

$$\mathbb{P}\{N_{t+s}-N_s=k\} = \mathbb{P}\{N_t=k\} = \frac{e^{-\lambda t}(\lambda t)^k}{k!} \text{ for } k=0,1,2,\cdots$$

(4) For $0\leq q<r\leq s<t$,

N_t-N_s and N_r-N_q are independent RVs

(but note $0\leq N_q\leq N_r\leq N_s\leq N_t$, so these RVs are <u>not</u> independent)

Rate of arrivals is λ arrivals per unit time:

$\mathbb{E}[N_t] = \lambda t$ because N_t follows Poisson distribution

<u>Review of Poisson distribution</u> $X \sim \text{Poisson}(\lambda)$

$$\mathbb{P}\{X = k\} = \frac{e^{-\lambda}\lambda^k}{k!}$$

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k \cdot \frac{e^{-\lambda}\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} e^{\lambda}$$

$$= \lambda$$

In R, dpois prob distribution $\mathbb{P}\{X=k\}$

ppois cumulation dist $\mathbb{P}\{X \le k\}$

qpois quantile

rpois random sample from Poisson distribution

Example: Suppose calls arrive on average 4 times per hour at an office phone. $\Rightarrow$ rate parameter $\lambda = 4/_{hr}$

(1) $\mathbb{P}\{$ fewer than 2 call during 1st hour $\} = \mathbb{P}\{N_1 = 0\} + \mathbb{P}\{N_1 = 1\}$

$$= e^{-4} + 4e^{-4}$$

$$= 5e^{-4} \approx 0.09$$

(2) $\mathbb{P}\{$ at least 2 calls during 2nd hour $\mid$ 6 calls during 1st hour $\}$

$= \mathbb{P}\{$ at least 2 calls during an hour $\}$ independence of nonoverlapping intervals

$= \mathbb{P}\{N_1 \geq 2\} = 1 - \mathbb{P}\{N_1 < 2\}$

$= 1 - 5e^{-4} \approx 0.91$

(3) $\mathbb{P}\{N_1 = 5 \mid N_2 = 8\} = \dfrac{\mathbb{P}\{5 \text{ calls in } 1^{st} \text{ hr } \underline{\text{and}} \text{ 3 calls in } 2^{nd} \text{ hr}\}}{\mathbb{P}\{8 \text{ calls in 2 hrs}\}}$

(overlapping intervals)

$$= \frac{\mathbb{P}\{N_1 = 5\}\,\mathbb{P}\{N_1 = 3\}}{\mathbb{P}\{N_2 = 8\}} = \binom{8}{5} / 2^8$$

Section 6.2 Arrival and interarrival times

Let X_k = interarrival time for k^{th} arrival

(how long after previous arrival $(k-1)^{th}$ it takes for the next arrival k^{th})

$\mathbb{P}\{X_1 > t\} = \mathbb{P}\{N_t = 0\} = e^{-\lambda t}$, so X_k follows the exponential distr. with λ

(independence of nonoverlapping intervals)

Review of exponential distribution with parameter λ

pdf $f(t) = \lambda e^{-\lambda t}$

cdf $F(t) = \mathbb{P}\{X \le t\} = \int_0^t \lambda e^{-\lambda x} dx = 1 - e^{-\lambda t}$

$\mathbb{E}[X] = 1/\lambda$

$Var[X] = 1/\lambda^2$

Memoryless: $\mathbb{P}\{X > s+t \mid X > s\} = \dfrac{e^{\lambda(s+t)}}{e^{\lambda s}} = e^{\lambda t} = \mathbb{P}\{X > t\} \quad \forall s,t>0$

(waiting starts over)

Alternate def'n of Poisson process (using interarrival times)

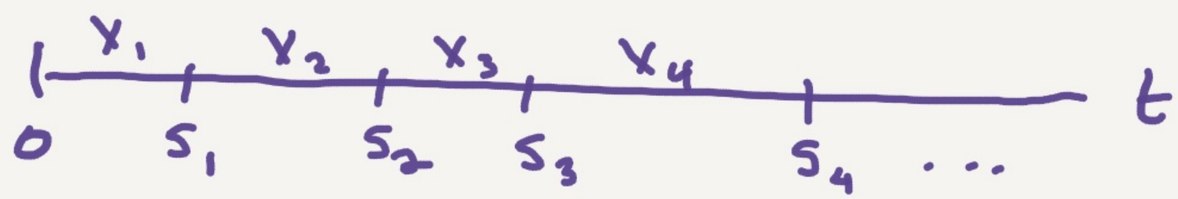
Let $X_1, X_2, \cdots$ be a sequence of iid $\exp(\lambda)$ RVs

For $t > 0$, let $N_t = \max\{n : X_1 + X_2 + \cdots + X_n \le t\}$ where $N_0 = 0$.

Then $(N_t)_{t \ge 0}$ defines a Poisson process with parameter λ.

Let $S_n = X_1 + \cdots + X_n$ for $n = 1, 2, \cdots$ (arrival time of nth customer)

Note $X_k = S_k - S_{k-1}$ (interarrival time between $(k-1)^{th}$ & k^{th} customers)



$\boxed{\text{R demo}}$

Given Poisson process $(N_t)_{t \geq 0}$ with rate parameter $\lambda = 4$

(average 4 customers per hr)

(1) $\mathbb{P}\{$wait at least 1 hr for 1st customer$\} = \mathbb{P}\{X_1 \geq 1\} = e^{-4} \approx 0.02$

(2) $\mathbb{P}\{$wait less than half an hr for 1st customer$\} = \mathbb{P}\{X_1 < \frac{1}{2}\}$

$$= 1 - \mathbb{P}\{X_1 \geq \tfrac{1}{2}\}$$

$$= 1 - e^{-4 \cdot \frac{1}{2}} \approx 0.86$$

(3) $\mathbb{P}\{$no customers during 1st hr | no customers during 1st 20 min$\}$

$= \mathbb{P}\{$takes 40 more minutes for 1st call$\}$

(like restarting process after 20 min)

$= \mathbb{P}\{X_1 > \frac{40}{60}\} = e^{-4 \cdot \frac{2}{3}} \approx 0.07$

Racing processes: minimum of independent exponential RVs

(who arrives first?)

Let $X_1, X_2, \cdots, X_n$ be independent exp RVs with parameters $\lambda_1, \lambda_2, \cdots, \lambda_n$

Let $M = \min\{X_1, \cdots, X_n\}$ (time for 1st arrival among the n competing Poisson processes)

Then $\mathbb{P}\{M > t\} = e^{-\lambda_1 t} e^{-\lambda_2 t} \cdots e^{-\lambda_n (t)} = e^{-(\lambda_1 + \lambda_2 + \cdots + \lambda_n)t}$

(they all take at least time t for 1st arrival)

Prob that k^{th} process wins (has 1st arrival before any of the other processes)

$$\mathbb{P}\{M = X_k\} = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_n}$$

Proof: (WLOG) $\mathbb{P}\{M = X_1\} = \mathbb{P}\{X_k \geq X_1 \text{ for } k = 2, \cdots, n\}$

$$= \int_0^\infty \mathbb{P}\{X_2 > t \mid X_1 = t\} \cdots \mathbb{P}\{X_n > t \mid X_1 = t\} \, \mathbb{P}\{X_1 = t\} \, dt$$

$$= \int_0^\infty e^{-(\lambda_2 + \cdots + \lambda_n)t} \cdot \lambda_1 e^{-\lambda_1 t} = \lambda_1 \int_0^\infty e^{-(\lambda_1 + \cdots + \lambda_n)t} \, dt$$

$$= \frac{\lambda_1}{\lambda_1 + \lambda_2 + \cdots + \lambda_n}$$

Further exploration: given independent Poisson processes

$(N_t^{(1)})_{t\geq 0}$ and $(N_t^{(2)})_{t\geq 0}$ with rate parameters λ_1 & λ_2, resp.

Then $N_t^{(1)} + N_t^{(2)}$ is a Poisson process with rate parameter $\lambda_1 + \lambda_2$.

Proof: $\mathbb{P}\{N_t^{(1)} + N_t^{(2)} = n\} = \sum_{k=0}^{n} \mathbb{P}\{N_t^{(1)} = k\}\mathbb{P}\{N_t^{(2)} = n-k\}$

$$= \sum_{k=0}^{n} \frac{e^{-\lambda_1 t}(\lambda_1 t)^k}{k!} \cdot \frac{e^{-\lambda_2 t}(\lambda_2 t)^{n-k}}{(n-k)!}$$

$$= \frac{t^n}{n!} e^{-t(\lambda_1+\lambda_2)} \sum_{k=0}^{n} \binom{n}{k} \lambda_1^k \lambda_2^{n-k}$$

apply binomial theorem

$$= \frac{t^n}{n!} e^{-t(\lambda_1+\lambda_2)} (\lambda_1+\lambda_2)^n$$

$$= \frac{e^{-t(\lambda_1+\lambda_2)}((\lambda_1+\lambda_2)t)^n}{n!}$$

So $N_t^{(1)} + N_t^{(2)}$ follows Poisson$((\lambda_1+\lambda_2)t)$ distribution

Also have $N_0^{(1)} + N_0^{(2)} = 0$,

$N_{t+s}^{(1)} + N_{t+s}^{(2)} - N_s^{(1)} - N_s^{(2)}$

has same distribution as

$N_t^{(1)} + N_t^{(2)}$ for all $t, s > 0$,

and for $0 \leq q < r \leq s < t$,

$N_t^{(1)} + N_t^{(2)} - N_s^{(1)} - N_s^{(2)}$ and

$N_r^{(1)} + N_r^{(2)} - N_q^{(1)} - N_q^{(2)}$

are independent

(nonoverlapping intervals, inherit properties from $N_t^{(1)}$ & $N_t^{(2)}$)

Example: Two coffee shops on a block with same rate parameter λ.
What is the probability that shop #1 receives 1st customer?

$$\frac{\lambda}{\lambda+\lambda} = 1/2$$

Now suppose shop #2 attracts 3 times as many customers.
What is probability that shop #2 receives 1st customer?

$$\frac{3\lambda}{\lambda+3\lambda} = 3/4$$

Example: Given two shops with resp. arrival rates of $\lambda_1 = 3$ customers/hr, $\lambda_2 = 6$/hr

(1) Probability that the 1st customer arrives at shop #1

$$\frac{\lambda_1}{\lambda_1+\lambda_2} = \frac{3}{9} = \frac{1}{3}$$

(2) Probability of 4 total customers in 1st half hr at the two shops

$$N_t = N_t^{(1)} + N_t^{(2)} \text{ with } \lambda = 3+6 = 9 \Rightarrow \mathbb{P}\{N_{1/2} = 4\} = \frac{e^{-9/2}(9/2)^4}{4!} \approx 0.19$$

(3) Given 4 total customers in 1st half hr, what is probability all 4 went to shop #2?

$$\mathbb{P}\{N_{1/2}^{(2)} = 4 \mid N_{1/2}^{(1)} + N_{1/2}^{(2)} = 4\} = \frac{\mathbb{P}\{N_{1/2}^{(1)} = 0\}\,\mathbb{P}\{N_{1/2}^{(2)} = 4\}}{\mathbb{P}\{N_{1/2}^{(1)} + N_{1/2}^{(2)} = 4\}} \approx .20 \text{ or } \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^4$$

does ${}^4C_4$ logic apply here?

Example: Suppose Amir, Beth, & Cathy arrive together at the post office, where 2 counters are open for service.

Assume service times are independent with exp(5 customers/hr)

Amir & Beth are immediately served.
Cathy will go to whichever counter is first free.

(1) Expected service time for each person: $1/\lambda = 1/5$ hr $= 12$ minutes

(2) How long should Cathy expect to wait on average?

Treat as sum of Poisson processes with $\lambda = 10 \Rightarrow 1/10$ hr $= 6$ minutes

(3) Total expected time for Cathy to be at post office:
(her wait time plus service time) Wait time + service time $= 6 + 12 = 18$ min.

(4) What is the probability that Cathy is the last to leave? $1/2$ due to memoryless property ?

(5) What is expected wait time before last person leaves?

$$\underbrace{\frac{1}{10}}_{\text{Cathy waiting}} + \underbrace{\frac{1}{10}}_{\text{Remaining 2 people}} + \underbrace{\frac{1}{5}}_{\text{Last person}} = \frac{2}{5}\text{ hr} = 24 \text{ min.}$$

Arrival times and the gamma distribution

Let S_n = time of nth arrival in a Poisson process with rate parameter λ

Then S_n follows the gamma distribution $f(t) = \dfrac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}$ for $t > 0$

$\mathbb{E}[S_n] = n/\lambda$

$Var[S_n] = n/\lambda^2$

(Note: $n=1$ case reduces to $\exp(\lambda)$)

Return to post office example: Suppose only 1 counter is open.

How long on average for Amir, Beth, & Cathy to be served?

$n = 3$ customers

$\lambda = 5$ customers/hr

$\mathbb{E}[S_3] = 3/5$ hr = 36 minutes

in R, cdf for S_n is

pgamma(t, n, λ)

(shape = n, rate = λ)

Derivation: want to show $f(t) = \dfrac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}$ for $t>0$

is the density function for S_n, $n=1,2,3,\cdots$

Base case: $S_1 = X_1$ has density $f(t) = \lambda e^{-\lambda t}$ (exponential distribution)

Induction step: assume $f_n(t) = \dfrac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}$ is density fn for S_n

Then density for $\underset{S_n + X_{n+1}}{\underset{\shortparallel}{S_{n+1}}}$ is $f_{n+1}(t) = \int_0^t \underbrace{\dfrac{\lambda^n t_1^{n-1} e^{-\lambda t_1}}{(n-1)!}}_{\text{first } n \text{ arrivals occur in time } t_1} \cdot \underbrace{\lambda e^{-\lambda(t-t_1)}}_{(n+1)^{st} \text{ arrival takes remaining time } t-t_1} dt_1$

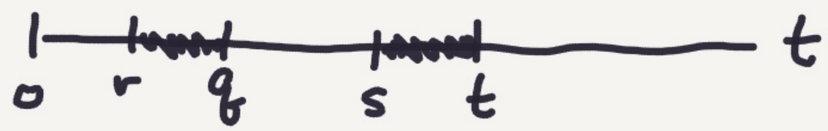
$$= \frac{\lambda^{n+1}}{(n-1)!} e^{-\lambda t} \int_0^t t_1^{n-1} dt_1$$

$$= \frac{\lambda^{n+1}}{(n-1)!} e^{-\lambda t} \cdot \frac{t^n}{n}$$

$$= \frac{\lambda^{n+1} t^n e^{-\lambda t}}{n!}$$

6.3 Infinitesimal probabilities : 3rd way to define Poisson process

Basic assumptions : (1) # of arrivals during one time interval does not affect # of arrivals in nonoverlapping time interval



(2) Assume average rate λ at which arrivals occur is constant over time

(3) Arrivals occur one at a time (never 2 simultaneously)

Consider what happens during a very short time interval $[t, t+\Delta t]$

Let N_t = # arrivals during $[0, t]$.

How many more arrivals occur during $(t, t+\Delta t]$?

(1) says # arrivals during $[0,t]$ and $(t, t+\Delta t]$ are independent

(2) says we expect λt arrivals on average during $[0, t]$ and
$\lambda \Delta t$ " during $(t, t+\Delta t]$.

(3) implies more than one arrival during $(t, t+\Delta t]$ is unlikely as $\Delta t \to 0$

More precisely, we can define Poisson process $(N_t)_{t \geq 0}$ as follows

(1) $N_0 = 0$

(2) $\underbrace{N_{t_1} - N_{s_1}}_{\#\text{ arrivals during } [s_1, t_1]}$ and $\underbrace{N_{t_2} - N_{s_2}}_{\#\text{ arrivals during } [s_2, t_2]}$ are independent if $0 \leq s_1 < t_1 \leq s_2 < t_2$

(3) $\mathbb{P}\{N_{t+\Delta t} = N_t\} = 1 - \lambda \Delta t + \underbrace{o(\Delta t)}_{\text{goes to } 0 \text{ faster than } \Delta t \text{ as } \Delta t \to 0}$

(4) $\mathbb{P}\{N_{t+\Delta t} = N_t + 1\} = \lambda \Delta t + o(\Delta t)$ (prob of 1 arrival during $(t, t+\Delta t]$)

(5) $\mathbb{P}\{N_{t+\Delta t} \geq N_t + 2\} = o(\Delta t)$ (highly unlikely for very small Δt)

Using these assumptions, we want to show the distribution of N_t agrees with our original definition of Poisson process:

$$\mathbb{P}\{N_t = k\} = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

Let $P_k(t) = \mathbb{P}\{N_t = k\}$. Always assume $N_0 = 0$.

$P_0(0) = 1$ and $P_k(0) = 0$ for $k = 1, 2, 3, \cdots$

Calculate $\mathbb{P}\{N_{t+\Delta t} = k\} = \mathbb{P}\{N_t = k\}\mathbb{P}\{N_{t+\Delta t} = k \mid N_t = k\}$ 0 arrivals during $(t, t+\Delta t)$

$+ \mathbb{P}\{N_t = k-1\}\,\mathbb{P}\{N_{t+\Delta t} = k \mid N_t = k-1\}$ 1 arrival during $(t, t+\Delta t)$

$+ \mathbb{P}\{N_t \leq k-2\}\,\mathbb{P}\{N_{t+\Delta t} = k \mid N_t \leq k-2\}$ ≥ 2 arrivals "

$$= P_k(t)\cdot(1 - \lambda\Delta t + o(\Delta t))$$
$$+ P_{k-1}(t)\cdot(\lambda\Delta t + o(\Delta t))$$
$$+ (P_0(t) + \cdots + P_{k-2}(t))\cdot o(\Delta t)$$
$$= P_k(t)(1 - \lambda\Delta t) + P_{k-1}(t)\cdot\lambda\Delta t + o(\Delta t)$$

coefficient for $o(\Delta t)$ is $P_0(t) + P_1(t) + \cdots P_k(t) \leq 1$

Take derivative to obtain a differential equation:

$$P_k'(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t}(\mathbb{P}\{N_{t+\Delta t} = k\} - \mathbb{P}\{N_t = k\}$$ usual difference quotient limit

$$= \lim_{\Delta t \to 0} \frac{1}{\Delta t}\left(P_k(t)\cdot(\cancel{1} - \lambda\Delta t) + P_{k-1}(t)\cdot\lambda\Delta t + o(\Delta t) - \cancel{P_k(t)}\right)$$

$$= \lim_{\Delta t \to 0}\left(-\lambda P_k(t) + \lambda P_{k-1}(t) + \frac{o(\Delta t)}{\Delta t}\right) \Rightarrow \boxed{P_k'(t) = \lambda P_{k-1}(t) - \lambda P_k(t)}$$

$$\boxed{P_k'(t) = \lambda P_{k-1}(t) - \lambda P_k(t)}$$ for $k = 0, 1, 2, \cdots$ and $t \geq 0$

$k=0$: $P_0'(t) = -\lambda P_0(t)$ and $P_0(0) = 1$, so $P_0(t) = e^{-\lambda t}$

Let $f_k(t) = e^{\lambda t} P_k(t)$, so $f_0(t) = 1$ and $f_k'(t) = \lambda f_{k-1}(t)$, $f_k(0) = 0$ for $k \geq 1$
(using boxed eq.)

Then for $k=1$, $f_1'(t) = \lambda$, $f_1(0) = 0 \Rightarrow f_1(t) = \lambda t$

$k=2$, $f_2'(t) = \lambda f_1(t) = \lambda^2 t$, $f_2(0) = 0 \Rightarrow f_2(t) = \frac{1}{2}\lambda^2 t^2$

$k=3$, $f_3'(t) = \lambda f_2(t) = \frac{1}{2}\lambda^3 t^2$, $f_3(0) = 0 \Rightarrow f_3(t) = \frac{1}{6}\lambda^3 t^3$

Conjecture $f_k(t) = \dfrac{\lambda^k t^k}{k!}$

Induction step to prove: suppose $f_k(t) = \dfrac{\lambda^k t^k}{k!}$

then $f_{k+1}'(t) = \lambda f_k(t) = \dfrac{\lambda^{k+1} t^k}{k!}$, $f_{k+1}(0) = 0$,

leading to $f_{k+1}(t) = \dfrac{\lambda^{k+1} t^{k+1}}{(k+1)!}$ ✓

Hence $P_k(t) = \dfrac{f_k(t)}{e^{\lambda t}} = \dfrac{e^{-\lambda t}(\lambda t)^k}{k!}$ and $(N_t)_{t \geq 0}$ follows the Poisson distribution.

Chapter 7 Continuous-time Markov chains

The Markov property is akin to the memoryless property of the Poisson process, associated with the exponential distribution:

Let $g(t) = \mathbb{P}\{X > t\}$.

To be <u>memoryless</u>, require $\mathbb{P}\{X > t\} = \mathbb{P}\{X > s+t \mid X > s\}$

That is, $\mathbb{P}\{X > t\} = \dfrac{\mathbb{P}\{X > s+t \ \& \ X > s\}}{\mathbb{P}\{X > s\}} = \dfrac{\mathbb{P}\{X > s+t\}}{\mathbb{P}\{X > s\}}$

$\Leftrightarrow \quad g(t)\,g(s) = g(s+t)$. What functions $g(t)$ satisfy this property?

More generally, we have $g(t_1) \cdots g(t_n) = g(t_1 + \cdots + t_n)$.

Take any $\frac{p}{q} \in \mathbb{Q}$. $g\left(\frac{p}{q}\right) = g\left(\frac{1}{q} + \overset{p \text{ copies}}{\cdots} + \frac{1}{q}\right) = g\left(\frac{1}{q}\right)^p$

$$g(1) = g\left(\frac{1}{q} + \overset{q \text{ copies}}{\cdots} + \frac{1}{q}\right) = g(1/q)^q$$

Hence $g\left(\frac{p}{q}\right) = g(1)^{p/q}$. For any $x \in \mathbb{R}$, use limit of rationals to obtain

$$g(x) = g(1)^x.$$

We now have $g(x) = g(1)^x$ for all $x \in \mathbb{R}$

$$= e^{\ln(g(1)) \cdot x}$$

$$= e^{-\lambda x}, \text{ where } \lambda = -\ln(g(1))$$

note $g(1) = \mathbb{P}\{X > 1\} \leq 1$
so $\ln(g(1)) \leq 0$

Therefore, for the RV X to satisfy the memoryless property, we require

$$\mathbb{P}\{X > t\} = e^{-\lambda t} \text{ for some } \lambda > 0, \text{ that is, } X \sim \text{Exp}(\lambda).$$

We want to build on some of the Poisson process & discrete Markov chain ideas to create a model for a continuous-time Markov chain.

In particular, what will replace the one-step transition matrix P?

A few definitions to start with:

Def'n: A continuous-time stochastic process $(X_t)_{t\geq 0}$ with a discrete state space S is a <u>continuous-time Markov chain</u> if

Markov property
$$\mathbb{P}\{X_{t+s}=j \mid X_s=i,\ X_u=x_u \text{ for } 0\leq u<s\} = \mathbb{P}\{X_{t+s}=j \mid X_s=i\}$$
for all $s,t\geq 0$, $i,j,x_u\in S$.

The process is <u>time-homogeneous</u> if this probability does not depend on s:
$$\mathbb{P}\{X_{t+s}=j \mid X_s=i\} = \mathbb{P}\{X_t=j \mid X_0=i\} \text{ for all } s,t\geq 0,\ i,j\in S.$$

The <u>transition function</u> is $P_{ij}(t) = \mathbb{P}\{X_t=j \mid X_0=i\}$

($P_{ij}(t)$ replaces the n-step transition matrix P^n_{ij} in the discrete-time MC)

Chapman-Kolmogorov Equations:

$$P(s+t) = P(s)P(t) \text{ for } s,t \geq 0$$

$$P_{ij}(s+t) = [P(s)P(t)]_{ij} = \sum_{k \in S} P_{ik}(s)P_{kj}(t) \quad \text{for all } i,j \in S,\ s,t \geq 0$$

Proof: $P_{ij}(s+t) = \mathbb{P}\{X_{s+t} = j \mid X_0 = i\}$ (def'n)

$$= \sum_{k \in S} \mathbb{P}\{X_{s+t} = j \mid X_s = k, \underbrace{X_0 = i}_{\text{doesn't affect prob by Markov property}}\} \mathbb{P}\{X_s = k \mid X_0 = i\}$$

$$= \sum_{k \in S} \mathbb{P}\{X_t = j \mid X_0 = k\} \mathbb{P}\{X_s = k \mid X_0 = i\} \quad \text{by time-homogeneity}$$

$$= \sum_{k \in S} P_{kj}(t) P_{ik}(s)$$

$$= [P(s)P(t)]_{ij} \quad \text{for all } i,j \in S,\ s,t \geq 0$$

Therefore $P(s+t) = P(s)P(t)$.

Example: Poisson process

$$P_{ij}(t) = \mathbb{P}\{N_{t+s} = j \mid N_s = i\} = \mathbb{P}\{N_t = j-i\} = \frac{e^{-\lambda t}(\lambda t)^{j-i}}{(j-i)!}$$

Holding times: Let T_i be the length of time that a cont-time MC, which started at state i, stays at state i before transitioning to a new state. (inter arrival time for Poisson process)

T_i has an exponential distribution because time-homogeneous cont-time MCs are memoryless (see bottom of page 268 to top of 269 in Dobrow text)

Examples: Queueing models $(X_t)_{t\geq 0}$, X_t = # waiting in line at time t (counting those being served)

M/M/1 queue: arrivals follow Poisson process with rate λ served by a single server at exponential rate μ

M/M/k queue: k servers, exponential rate $\begin{cases} k\mu & \text{if at least } k \text{ in queue} \\ n\mu & \text{if } n<k \text{ in queue} \end{cases}$

M/M/∞ queue: serve everyone in line, rate $n\mu$, n = # in line

Section 7.2 Alarm clocks & transition rates

We will use rate parameters q_{ij} rather than direct probabilities for transitions from states i to j (q_{ij} are like λ in Poisson process)

Imagine being at state i, with all other states racing to finish first and become the next state, like a set of alarm clocks where the first one to go off wins. Just as with racing Poisson processes, the prob of the next state being j is

$$P_{ij} = \frac{q_{ij}}{\sum_{k \in S} q_{ik}}$$

We need to set up a way of analyzing such a system.

Section 7.3 Infinitesimal generator

The key properties of a continuous-time (time-homogeneous) MC $(X_t)_{t\geq 0}$ are

① Markov property

$$\mathbb{P}\{X_{t+s}=j \mid X_s=i,\ X_u=x_u \text{ for } 0\leq u<s\} \quad \forall t,s\geq 0,\ i,j,x_u\in S$$

② Time-homogeneity

$$\mathbb{P}\{X_{t+s}=j \mid X_s=i\} = \mathbb{P}\{X_t=j \mid X_0=i\} \quad \forall t,s\geq 0,\ i,j\in S$$

Assume X_t is governed by <u>transition rates</u> q_{ij}

Let $q_i = \sum_{j\neq i} q_{ij}$ = total rate for state i (holding time parameter)

Mimicking the 3rd def'n of Poisson process,

$$\mathbb{P}\{X_{t+\Delta t}=i \mid X_t=i\} = 1 - q_i\Delta t + o(\Delta t)$$

$$\mathbb{P}\{X_{t+\Delta t}=j \mid X_t=i\} = q_{ij}\Delta t + o(\Delta t) \quad \text{for } j\neq i$$

$$\mathbb{P}\{X_{t+\Delta t} = i \mid X_t = i\} = 1 - q_i \Delta t + o(\Delta t)$$

$$\mathbb{P}\{X_{t+\Delta t} = j \mid X_t = i\} = q_{ij} \Delta t + o(\Delta t) \quad \text{for } j \neq i$$

Let $p_i(t) = \mathbb{P}\{X_t = i\}$

$$p_i'(t) = \lim_{\Delta t \to 0} \frac{p_i(t+\Delta t) - p_i(t)}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{1}{\Delta t}\left(\mathbb{P}\{X_{t+\Delta t} = i\} - \mathbb{P}\{X_t = i\}\right)$$

$$= \lim_{\Delta t \to \infty} \frac{1}{\Delta t}\Big(\sum_{j \neq i} \mathbb{P}\{X_{t+\Delta t} = i \mid X_t = j\}\mathbb{P}\{X_t = j\} + \mathbb{P}\{X_{t+\Delta t} = i \mid X_t = i\}\mathbb{P}\{X_t = i\} - \mathbb{P}\{X_t = i\}\Big)$$

$$= \lim_{\Delta t \to \infty} \frac{1}{\Delta t}\Big(\sum_{j \neq i} q_{ji} \Delta t \cdot p_j(t) + (1 - q_i \Delta t) p_i(t) - p_i(t) + o(\Delta t)\Big)$$

$$= \sum_{j \neq i} q_{ji}(t) p_j(t) - q_i p_i(t) + \cancel{\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t}} \to 0$$

We need to solve this system of DEs w/ initial condition

$$p_i(0) = \mathbb{P}\{X_0 = i\} \quad \text{for } i \in S$$

Assume a finite # of states and express in matrix form:

$$Q = \begin{bmatrix} -q_1 & q_{12} & \cdots & q_{1n} \\ q_{21} & -q_2 & \cdots & q_{2n} \\ \vdots & & \ddots & \vdots \\ q_{n1} & \cdots & & -q_n \end{bmatrix}$$

diagonal entries $-q_i$

non-diagonal entries q_{ij}

Q is called the <u>generator matrix</u>

We need to solve $[p_1'(t) \cdots p_n'(t)] = [p_1(t) \cdots p_n(t)]\, Q$

$$\vec{p}\,'(t) = \vec{p}(t)\, Q$$

The solution to this linear system is $\vec{p}(t) = \vec{p}(0)\, e^{tQ}$

Hence the transition function is $P(t) = e^{tQ}$

We need background on the matrix exponential e^{tQ} to make sense of this solution.

Def'n: $e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!}$ (like $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$)

Basic properties:

① $e^0 = I$

② $e^A e^{-A} = I$

③ $e^{(s+t)A} = e^{sA} e^{tA}$

④ $\frac{d}{dt} e^{tA} = e^{tA} A = A e^{tA}$

$I = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$

$AI = IA = A$

Calculating powers of matrices is computationally intensive, so we apply diagonalization for a more efficient approach.

$Q = SDS^{-1}$, $D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$, $S = [\vec{v}_1 \cdots \vec{v}_n]$, $S^{-1} = \begin{bmatrix} \vec{w}_1 \\ \vdots \\ \vec{w}_n \end{bmatrix}$

right eigenvectors
$Q\vec{v}_k = \lambda_k \vec{v}_k$

left eigenvectors
$\vec{w}_k Q = \lambda_k \vec{w}_k$

$$P(t) = e^{tQ} = \sum_{k=0}^{\infty} \frac{t^k}{k!} (SDS^{-1})^k = S \sum_{k=0}^{\infty} \frac{t^k}{k!} \begin{bmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_n^k \end{bmatrix} S^{-1} = S \begin{bmatrix} \sum_k \frac{(\lambda_1 t)^k}{k!} & & \\ & \ddots & \\ & & \ddots \end{bmatrix} S^{-1}$$

$$= S \begin{bmatrix} e^{\lambda_1 t} & & \\ & \ddots & \\ & & e^{\lambda_n t} \end{bmatrix} S^{-1}$$

Verify that $\frac{d}{dt}\left[e^{tA}\right] = e^{tA}A$

$$\frac{d}{dt}\left[e^{tA}\right] = \sum_{n=0}^{\infty} \frac{d}{dt}\left[\frac{t^n A^n}{n!}\right] = \sum_{n=1}^{\infty} \frac{n t^{n-1} A^n}{n!}$$

$n=0$ term is constant, so derivative is 0

$$= \sum_{n=0}^{\infty} \frac{(n+1) t^n A^{n+1}}{(n+1)!} = \sum_{n=0}^{\infty} \frac{t^n A^n}{n!} \cdot A = e^{tA}A$$

(change of index to match defn of e^{tA})

Our diff eq is $\vec{p}\,'(t) = \vec{p}(t)Q$.

Let $\vec{p}(t) = \vec{p}(0)e^{tQ}$.

This satisfies the initial condition at $t=0$ since $e^{\overbrace{0Q}^{\text{zero matrix}}} = I$

and also satisfies the DE:

$$\vec{p}\,'(t) = \vec{p}(0)\frac{d}{dt}\left[e^{tQ}\right] = \underbrace{\vec{p}(0)e^{tQ}}_{\vec{p}(t)}Q = \vec{p}(t)Q$$

Therefore e^{tQ} is the transition function we seek.

As with the discrete-time MC, multiplying by the initial distribution $\vec{p}(0)$ gives us a vector corresponding to the desired starting state.

Other linear algebra facts about Q:

Rows of Q sum to zero, so $\det(Q - 0I) = \det(Q) = 0$,
implying Q will always have an eigenvalue $\lambda = 0$
(taking place of $\lambda = 1$ for discrete-time transition matrix P).

Note that Q is a matrix of <u>rates</u>, not probabilities.

Diagonalization: why $S = [\vec{v}_1 \cdots \vec{v}_n]$ and $S^{-1} = \begin{bmatrix} \vec{w}_1 \\ \vdots \\ \vec{w}_n \end{bmatrix}$, $Q = SDS^{-1}$,
where $A\vec{v}_k = \lambda \vec{v}_k$ and $\vec{w}_k A = \lambda_k \vec{w}_k$, $D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$

$$QS = [Q\vec{v}_1 \cdots Q\vec{v}_n] = [\lambda_1 \vec{v}_1 \cdots \lambda_n \vec{v}_n] = [\vec{v}_1 \cdots \vec{v}_n] \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = SD$$

so $Q = SDS^{-1}$.

Let $R = [\vec{w}_1 \cdots \vec{w}_n]$. $RQ = \begin{bmatrix} \vec{w}_1 \\ \vdots \\ \vec{w}_n \end{bmatrix} Q = \begin{bmatrix} \vec{w}_1 Q \\ \vdots \\ \vec{w}_n Q \end{bmatrix} = \begin{bmatrix} \lambda_1 \vec{w}_1 \\ \vdots \\ \lambda_n \vec{w}_1 \end{bmatrix} = DR$

Hence $Q = SDS^{-1} = R^{-1}QR$. Diagonalization is unique if fix order and normalization of eigenvectors, implying $S^{-1} = R$.

Example: $S = \{1, 2, 3\}$

$\boxed{\text{R demo}}$

$$Q = \begin{bmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{bmatrix}$$

note rows sum to 0

Key eigenvalue

Eigenvalues $\lambda = 0, -3, -3$

Right eigenvectors $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$

Left eigenvectors $\begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix}$

$\begin{bmatrix} -1/3 & 1/6 & 1/6 \end{bmatrix}$

$\begin{bmatrix} 0 & -1/2 & 1/2 \end{bmatrix}$

$$P(t) = \begin{bmatrix} 1 & -2 & 0 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & e^{-3t} & \\ & & e^{-3t} \end{bmatrix} \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ -1/3 & 1/6 & 1/6 \\ 0 & -1/2 & 1/2 \end{bmatrix}$$

$$= \begin{bmatrix} 1/3 + 2/3\, e^{-3t} & 1/3 - 1/3\, e^{-3t} & 1/3 - 1/3\, e^{-3t} \\ 1/3 - 1/3\, e^{-3t} & 1/3 + 2/3\, e^{-3t} & 1/3 - 1/3\, e^{-3t} \\ 1/3 - 1/3\, e^{-3t} & 1/3 - 1/3\, e^{-3t} & 1/3 + 2/3\, e^{-3t} \end{bmatrix}$$

$$\lim_{t \to \infty} P(t) = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} = \begin{bmatrix} \pi \\ \pi \\ \pi \end{bmatrix}$$

Recap of continuous-time Markov chain on discrete state space

- Markov property: only most recent info matters
- Time-homogeneous: rates are constant through time

$\Rightarrow$ once you know what state MC is in at the most recent time, it's like the chain starts over again from that state.

Transition function $P_{ij}(t) = \mathbb{P}\{X_t = j \mid X_0 = i\}$

$P(t) = e^{tQ}$, where Q is the generator matrix

$$Q = \begin{bmatrix} -q_1 & q_{12} & \cdots & q_{1n} \\ q_{21} & -q_2 & \cdots & q_{2n} \\ \vdots & & \ddots & \\ q_{n1} & & & -q_n \end{bmatrix}$$

q_{ij} is the transition rate from state i to j

$q_i = \sum_{j \neq i} q_{ij}$

If currently in state i, prob that next state is j equals

$$\frac{q_{ij}}{\sum_{k \neq i} q_{ik}}$$

(racing processes)

7.4 Long-term behavior

Def'n: A probability distribution π is the <u>limiting distribution</u> of a cont-time MC if for all states i, j we have

$$\lim_{t\to\infty} P_{ij}(t) = \pi_j \quad \left(\text{equivalently, } \lim_{t\to\infty} P(t) = \begin{bmatrix} \pi \\ \vdots \\ \pi \end{bmatrix}\right)$$

Def'n: A probability distribution π is a <u>stationary distribution</u> if $\pi = \pi P(t)$ for all $t \geq 0$, that is,

(left eigenvector of Q for $\lambda = 0$)

$$\pi_j = \sum_i \pi_i P_{ij}(t) \text{ for all states } j \text{ and time } t \geq 0.$$

* Limiting distr. is always stationary, but not vice versa.

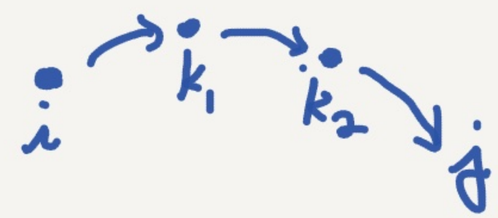
Def'n: A cont-time MC is <u>irreducible</u> if for every i & j there exists $t > 0$ with $P_{ij}(t) > 0$

(can eventually get from every state to every other state)

Lemma 7.1 If $P_{ij}(t) > 0$ for some $t > 0$, then $P_{ij}(t) > 0$ for all $t > 0$.

(This is very different from the discrete time case.)

Intuition: process is rate-based, so can have very short transition times (or very long ones). Even if need to hit intermediate states to get from i to j, it's possible to transition through them very quickly (or very slowly).



Defn: A state i is <u>absorbing</u> if $q_i = \sum_{j \neq i} q_{ij} = 0$

If all rates q_{ij} and q_i are finite, then either

(1) The MC is irreducible, all states communicate, $P_{ij}(t) > 0$ for all states i & j and time $t > 0$.

OR

(2) The MC has at least one absorbing state.

(q_i could be infinite in case of infinite state space)

Theorem 7.2: Fundamental Limit Theorem

Let $(X_t)_{t\geq 0}$ be a finite, irreducible, cont-time, time-homogeneous MC with transition fn $P(t)$. Then there exists a unique stationary distribution $\bar{\pi}$, which is also the limiting distribution.

For all states i,j, $\lim_{t\to\infty} P_{ij}(t) = \pi_j$, $\lim_{t\to\infty} P(t) = \begin{bmatrix} \bar{\pi} \\ \vdots \\ \bar{\pi} \end{bmatrix}$, and $\bar{\pi} = \bar{\pi} P(t)$, $t \geq 0$.

Recall $P(t) = e^{tQ}$ and $P'(t) = P(t)Q = QP(t)$.

At time $t=0$, $P'(0) = P(0)Q = IQ = Q$.

For a stationary distr $\bar{\pi}$, we have $\bar{\pi} = \bar{\pi} P(t)$. Taking a derivative of both sides,

$$0 = \bar{\pi} P'(t) = \bar{\pi} P(t) Q = \bar{\pi} Q$$

On the other hand, if $\bar{\pi}$ satisfies $0 = \bar{\pi} Q$, then $0 = \bar{\pi} Q P(t) = \bar{\pi} P'(t)$.

Because $P(0) = I$, so $\bar{\pi} P(0) = \bar{\pi}$, we must have $\bar{\pi} = \bar{\pi} P(t)$ for all $t \geq 0$.

Therefore, being a stationary distr $\bar{\pi}$ is equivalent to $0 = \bar{\pi} Q$, implying that $\bar{\pi}$ is a left eigenvector of Q for eigenvalue $\lambda = 0$.

Stationary distributions and the generator matrix Q:

A probability distribution π is a stationary distribution of a cont-time MC with generator matrix Q if and only if

$\pi Q = 0$, that is, $\sum_i \pi_i Q_{ij} = 0$ for all states j.

Example: Suppose a puppy is always in one of 3 states: eat (1), play (2), sleep (3).

Average time eating a meal is 15 min = 1/4 hr $\rightarrow q_1 = 4/\text{hr}$

play bout 30 min = 1/2 hr $\rightarrow q_2 = 2$

nap 60 min = 1 hr $\rightarrow q_3 = 1$

After eating, 1/2 prob of play, 1/2 of nap $\rightarrow q_{12} = 2 \quad q_{13} = 2$

playing, 1/3 prob of eat, 2/3 of nap $\rightarrow q_{21} = 2/3 \quad q_{23} = 4/3$

napping, 1/4 prob of eat, 3/4 of play $\rightarrow q_{31} = 1/4 \quad q_{32} = 3/4$

$\boxed{\text{Use } q_{ij} = q_i P_{ij}}$

$\boxed{\text{R demo}}$

$$Q = \begin{array}{c} \\ \text{eat} \\ \text{play} \\ \text{nap} \end{array} \begin{array}{c} \begin{array}{ccc} \text{eat} & \text{play} & \text{nap} \end{array} \\ \begin{bmatrix} -4 & 2 & 2 \\ 2/3 & -2 & 4/3 \\ 1/4 & 3/4 & -1 \end{bmatrix} \end{array}$$

note rows sum to zero

$$\pi = \begin{array}{c} \begin{array}{ccc} \text{eat} & \text{play} & \text{nap} \end{array} \\ [.09 \quad .31 \quad .60] \end{array}$$

Absorbing cont-time Markov chains

Let $(X_t)_{t\geq 0}$ be a cont-time, time-homogeneous MC with finite state space. Suppose it has exactly one absorbing state a with positive probability $P_{ia}(t)$ for some time $t>0$ for all other states i. That is, all states are transient except for the absorbing state a.

What is the expected time to absorption?

Rearrange Q to have a in the upper left corner:

$$Q = \begin{array}{c|c} & \begin{array}{cccc} a & i_1 & \cdots & i_{n-1} \end{array} \\ \hline \begin{array}{c} a \\ i_1 \\ \vdots \\ i_{n-1} \end{array} & \begin{bmatrix} 0 & 0 & \cdots & 0 \\ * & & & \\ \vdots & & V & \\ * & & & \end{bmatrix} \end{array}$$

transient states

rates between transient states

Defn: Fundamental matrix $\boxed{F = -V^{-1}}$

F will provide a way to calculate the <u>mean time until absorption</u> for a chain started at a transient state i: $a_i = \mathbb{E}[\text{first time } X_t = a \mid X_0 = i]$

Consider the exponentially distributed waiting times for X_t:

let $T_i = \min\{t>0 : X_t \neq i \mid X_0 = i\}$ (first time leave state i)

$\mathbb{P}\{X_{T_i} = j \mid X_0 = i\} = \frac{q_{ij}}{q_i}$ (prob next state is j) and $\mathbb{E}[T_i] = \frac{1}{q_i}$

Next consider $Y = \min\{t>0 : X_t = a\}$ (first time hit absorbing state a)

$$a_i = \mathbb{E}[Y \mid X_0 = i] = \underbrace{\mathbb{E}[T_i]}_{\text{time to leave state } i} + \sum_{k \neq i,a} \underbrace{\mathbb{P}\{X_{T_i} = k \mid X_0 = i\}}_{\text{at next state } k \text{ (neither } i \text{ nor } a)} \underbrace{\mathbb{E}[Y \mid X_0 = k]}_{\text{time to wait now that at state } k}$$

$$= \frac{1}{q_i} + \sum_{k \neq i,a} \frac{q_{ik}}{q_i} a_k$$

$$a_i q_i = 1 + \sum_{k \neq i,a} q_{ik} a_k$$ (some rearranging)

$$-q_i a_i + \sum_{k \neq i,a} q_{ik} a_k = -1$$

$$V \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} -q_1 & q_{12} & \cdots & q_{1n} \\ q_{21} & -q_2 & \cdots & \\ & & \ddots & \\ & & & -q_n \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} -1 \\ \vdots \\ -1 \end{bmatrix} \Rightarrow \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = -V^{-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \underbrace{F \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}}_{\text{summing across each row of } F}$$

(over n transient states)

Theorem 7.3 For a cont-time MC with one absorbing state a and started at state $i \neq a$, the mean time until absorption into state a is $a_i = \sum_k F_{ik}$ (summing over transient states k)

To calculate the mean passage time from state i to state j in an irreducible cont-time MC, treat state j as absorbing.

$\boxed{\text{R demo}}$

$$Q = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} -3 & 1 & 1 & 1 \\ 1 & -2 & 0 & 1 \\ 2 & 0 & -4 & 2 \\ 1 & 1 & 1 & -3 \end{bmatrix} \end{array}$$

$\bar{\pi} = [\,2/7 \;\; 2/7 \;\; 1/7 \;\; 2/7\,]$

$\mathbb{E}[T_1] = 1/3$ $\qquad$ $\mathbb{E}[T_3] = 1/4$

$\mathbb{P}\{\text{next state is } 2 \mid X_0 = 1\} = 1/3$

$\mathbb{P}\{\text{next state is } 4 \mid X_0 = 3\} = 1/2$

Mean passage time from state 2 to state 1:

$$V = \begin{array}{c} \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccc} 2 & 3 & 4 \end{array} \\ \begin{bmatrix} -2 & 0 & 1 \\ 0 & -4 & 2 \\ 1 & 1 & -3 \end{bmatrix} \end{array} \qquad F = -V^{-1} = \begin{array}{c} \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccc} 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 5/8 & 1/16 & 1/4 \\ 1/8 & 5/16 & 1/4 \\ 1/4 & 1/8 & 1/2 \end{bmatrix} \end{array}$$

$$\Rightarrow a_2 = \frac{5}{8} + \frac{1}{16} + \frac{1}{4} = \frac{15}{16} = 0.9375$$

Remark: discrete MC vs cont-time MC

The <u>embedded chain</u> $(\tilde{X}_n)_{n\geq 0}$ for cont-time MC $(X_t)_{t\geq 0}$ is the discrete-time chain that tracks the sequence of states but ignores the actual passage times (treat them as a single step)

For a cont-time MC, $\bar{\pi}$ satisfies $\bar{\pi} = \bar{\pi} P(t)$ and $\bar{\pi} Q = 0$, (left eigenvector for $\lambda = 0$ of Q) giving the long-term proportion of <u>time</u> spent in each state.

For a discrete-time MC, $\bar{\pi}$ satisfies $\bar{\pi} = \bar{\pi} P$, (left eigenvector for $\lambda = 1$ of P) giving the long-term proportion of <u>visits</u> to each state.

Returning to previous example, transition matrix $\tilde{P}$ for the embedded discrete-time MC has $\tilde{P}_{ij} = \frac{q_{ij}}{q_i}$ for $i \neq j$ with $\tilde{P} = \begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 & 0 \end{bmatrix}$

Q's $\bar{\pi} = [2/7 \;\; 2/7 \;\; 1/7 \;\; 2/7]$ proportion time spent in each state

$\tilde{P}$'s $\bar{\pi} = [.3 \;\; .2 \;\; .2 \;\; .3]$ proportion of visits to each state

Section 7.5 Time reversibility – birth & death chains

Def'n: A continuous-time MC with generator matrix Q and unique stationary distribution $\bar{\pi}$ is called <u>time-reversible</u> if

$$\pi_i q_{ij} = \pi_j q_{ji} \quad \text{for all states } i \text{ \& } j$$

"local balance eq's"

If you know a chain is reversible, you can use these local balance equations to find $\bar{\pi}$:

$$\sum_i \pi_i q_{ij} = \sum_i \pi_j q_{ji} = \pi_j \underbrace{\sum_i q_{ji}}_{\text{sum across row } j \text{ is } 0} = 0, \text{ so } \bar{\pi} Q = 0.$$

Birth and death process

- type of time-reversible, cont-time MC on state space $\{0, 1, 2, \cdots\}$
- births are transitions from i to $i+1$ at rate λ_i for $i \geq 0$
- deaths are transitions from i to $i-1$ at rate μ_i for $i \geq 1$

Generator matrix
$$Q = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \cdots \end{array} \\ \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & \cdots \\ \mu_1 & -\lambda_1 - \mu_1 & \lambda_1 & 0 \cdots \\ 0 & \mu_2 & -\lambda_2 - \mu_2 & \lambda_2 \\ & & & \end{bmatrix} \end{array}$$

$q_{i,i+1} = \lambda_i$

$q_{i,i-1} = \mu_i$

$q_i = \lambda_i + \mu_i$

$(q_{ij} = 0 \text{ for } j \neq i \pm 1)$

Local balance eq's
$$\pi_i q_{i,i+1} = \pi_{i+1} q_{i+1,i} \quad \text{for } i = 0, 1, 2, \cdots$$
$$\pi_i \lambda_i = \pi_{i+1} \mu_{i+1}$$

Solve:
$$\pi_1 = \pi_0 \frac{\lambda_0}{\mu_1}$$
$$\pi_2 = \pi_1 \frac{\lambda_1}{\mu_2} = \pi_0 \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2}$$

etc

Generally,
$$\pi_k = \pi_0 \frac{\lambda_0 \cdots \lambda_{k-1}}{\mu_1 \cdots \mu_k}$$

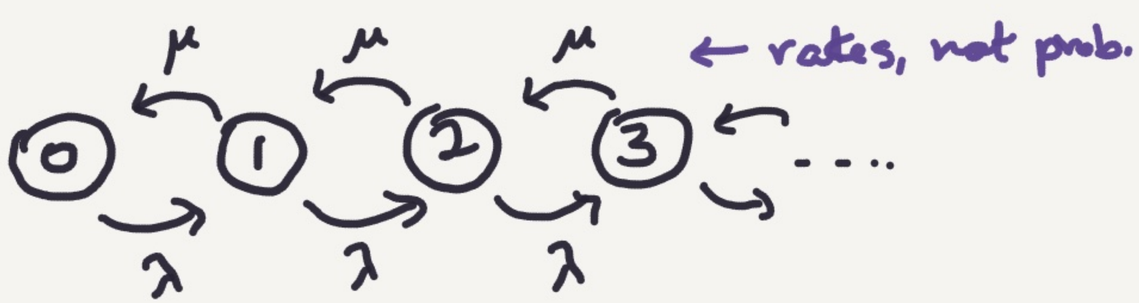
Birth and death process

$$\pi_k = \pi_0 \prod_{i=1}^{k} \frac{\lambda_{k-1}}{\mu_k} = \pi_0 \frac{\lambda_0 \cdots \lambda_{k-1}}{\mu_1 \cdots \mu_k}$$

We need $\sum_{k=0}^{\infty} \pi_k = 1$, so $\pi_0 = 1 \Big/ \underbrace{\sum_{k=0}^{\infty} \prod_{i=1}^{k} \frac{\lambda_{k-1}}{\mu_k}}$ $(k=0$ term is $1)$

This series must converge for a solution to exist.

Example: Cont-time random walk



← rates, not prob.

All birth rates $\lambda_i = \lambda$

All death rates $\mu_i = \mu$

$$\pi_k = \pi_0 \prod_{i=1}^{k} \frac{\lambda}{\mu} = \pi_0 \left(\frac{\lambda}{\mu}\right)^k \quad \text{for } k = 0, 1, 2, \cdots$$

$$\pi_0 = 1 \Big/ \underbrace{\sum_{k=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^k}_{\text{converges if } \frac{\lambda}{\mu} < 1} = 1 \Big/ \frac{1}{1 - \lambda/\mu} = 1 - \frac{\lambda}{\mu} \quad \text{if } \lambda < \mu$$

and so $\pi_k = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^k$

Section 7.6 Queueing theory

X_t = # waiting at time t (in line or being served)

Arrival rate λ to queue

Departure rate μ (service rate)

M/M/1 queue has a single server, so $\lambda_n = \lambda$ and $\mu_n = \mu$

M/M/k queue has k servers, so $\lambda_n = \lambda$

$$\mu_n = \begin{cases} n\mu & \text{if } n < k \\ k\mu & \text{if } n \geq k \end{cases}$$

M/M/∞ queue has everyone being served (infinite capacity),

so $\lambda_n = \lambda$

$\mu_n = n\mu$

$(X_t)_{t \geq 0}$ is irreducible here if all $\underset{(n \geq 0)}{\lambda_n}$ and $\underset{(n \geq 1)}{\mu_n}$ are positive

Note that an irreducible cont-time MC has all recurrent states if there is a probability of 1 of returning to each state (embedded discrete-time MC has all recurrent states),

while states are transient if prob < 1 of returning (meaning the queue tends to grow longer & longer over time)

The embedded discrete-time MC has $\tilde{P}_{i,i+1} = \frac{\lambda_i}{\lambda_i + \mu_i}$

$$\tilde{P}_{i,i-1} = \frac{\mu_i}{\lambda_i + \mu_i}$$

(otherwise $\tilde{P}_{ij} = 0$)

How do we tell if the queue is recurrent or transient?

Let $s(n) =$ prob of reaching state 0 eventually if start at state n $(X_0 = n)$

$s(0) = 1$ (already there)

$n \geq 1$: $s(n) = \mathbb{P}\{$next state is $n+1 \mid$ current state is $n\}\, s(n+1)$
(an arrival occurs next)

$+ \mathbb{P}\{$next state is $n-1 \mid$ current state is $n\}\, s(n-1)$
(a departure occurs next)

$$= \frac{\lambda_n}{\mu_n + \lambda_n} s(n+1) + \frac{\mu_n}{\mu_n + \lambda_n} s(n-1)$$

Multiply by $\mu_n + \lambda_n$: $(\mu_n + \lambda_n) s(n) = \lambda_n s(n+1) + \mu_n s(n-1)$

$$s(n) - s(n+1) = \frac{\mu_n}{\lambda_n}\left(s(n-1) - s(n)\right)$$

Iterate a few times to see pattern: $s(1) - s(2) = \frac{\mu_1}{\lambda_1}\left(s(0) - s(1)\right)$

$$s(2) - s(3) = \frac{\mu_2}{\lambda_2}\left(s(1) - s(2)\right) = \frac{\mu_1 \mu_2}{\lambda_1 \lambda_2}\left(s(0) - s(1)\right)$$

Telescoping sum: $s(n+1) = \sum_{j=0}^{n} \left(s(j+1) - s(j)\right) + s(0)$

So far: $s(n) - s(n+1) = \frac{\mu_1 \cdots \mu_n}{\lambda_1 \cdots \lambda_n} (\cancel{s(0)}^{\,1} - s(1))$

$$s(n+1) = s(0) + \sum_{j=0}^{n} (s(j+1) - s(j))$$

$$= 1 - \underbrace{\sum_{j=0}^{n} \frac{\mu_1 \cdots \mu_j}{\lambda_1 \cdots \lambda_j}}_{\text{interesting part}} \underbrace{(1 - s(1))}_{\text{constant}}$$

Recall: $s(n) = 1$ means guaranteed to return to state 0, so system is recurrent.

$s(n) < 1$ means transient.

Case $s(1) < 1$: assuming irreducible MC, so $s(n) < 1 \ \forall n \geq 0$

$$0 \leq \lim_{n \to \infty} s(n) = 1 - \sum_{j=0}^{\infty} \frac{\mu_1 \cdots \mu_j}{\lambda_1 \cdots \lambda_j} \underbrace{(1 - s(1))}_{>0} < 1, \text{ so } \sum_{j=0}^{\infty} \frac{\mu_1 \cdots \mu_j}{\lambda_1 \cdots \lambda_j} < \infty$$

<u>transient</u> equivalent to $\sum_{n=0}^{\infty} \frac{\mu_1 \cdots \mu_n}{\lambda_1 \cdots \lambda_n}$ converging

Cases: $s(1) = 1$ <u>recurrent</u>, equivalent to $\sum_{n=0}^{\infty} \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n}$ converging

(and $s(n) = 1 \ \forall n \geq 0$)

(see next slide for derivation of series)

Theorem A birth and death chain is:

transient iff $\sum_{n=1}^{\infty} \frac{\mu_1 \cdots \mu_n}{\lambda_1 \cdots \lambda_n} < \infty$

positive recurrent iff $\sum_{n=1}^{\infty} \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} < \infty$

null recurrent iff neither series converges

Note: to derive positive recurrent case, use a similar analysis but for π_n, where $\pi Q = 0$, requiring $\sum_{n=0}^{\infty} \pi_n = 1$

Solve $\mu_1 \pi_1 - \lambda_0 \pi_0 = 0$

$$\mu_{n+1} \pi_{n+1} + \lambda_{n-1} \pi_{n-1} - (\mu_n + \lambda_n) \pi_n = 0 \text{ for } n \geq 1$$

$$\Rightarrow \pi_1 = \frac{\lambda_0}{\mu_1} \pi_0$$

$$n=1: \pi_2 = \frac{1}{\mu_2}\left((\mu_1 + \lambda_1)\pi_1 - \lambda_0 \pi_0\right) = \frac{1}{\mu_2}\left((\mu_1 + \lambda_1)\frac{\lambda_0}{\mu_1}\pi_0 - \lambda_0 \pi_0\right)$$

$$= \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} \pi_0, \text{ etc}$$

Apply to queueing models:

<u>M/M/1 queue</u> $\lambda_n = \lambda$, $\mu_n = \mu$

$$\sum_{n=1}^{\infty} \frac{\mu_1 \cdots \mu_n}{\lambda_1 \cdots \lambda_n} = \sum_{n=1}^{\infty} \left(\frac{\mu}{\lambda}\right)^n = \frac{1}{1-\mu/\lambda} - 1 = \frac{\mu/\lambda}{1-\mu/\lambda}$$

converges if $\frac{\mu}{\lambda} < 1$

so transient if $\mu < \lambda$ (service slower than arrival rate)

$$\sum_{n=1}^{\infty} \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} = \sum_{n=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^n = \frac{\lambda/\mu}{1-\mu/\lambda}$$

converges if $\frac{\lambda}{\mu} < 1$

so positive recurrent if $\lambda < \mu$ (service faster than arrivals)

Queue is then null recurrent if $\lambda = \mu$.

(Will further explore in queueing lab.)

M/M/k queue $\quad \lambda_n = \lambda$

$$\mu_n = \begin{cases} n\mu & n < k \\ k\mu & n \geq k \end{cases}$$

For $n \geq k$, $\dfrac{\mu_1 \cdots \mu_n}{\lambda_1 \cdots \lambda_n} = \dfrac{\mu(2\mu)\cdots(k\mu)(k\mu)^{n-k}}{\lambda^n} = \dfrac{k!\,\mu^n k^{n-k}}{\lambda^n}$

$$= \frac{k!}{k^k}\left(\frac{k\mu}{\lambda}\right)^n$$

fixed value

$\Rightarrow$ series $\sum_{n=1}^{\infty} \dfrac{\mu_1 \cdots \mu_n}{\lambda_1 \cdots \lambda_n}$ converges if $k\mu < \lambda$

For $n \geq k$, $\dfrac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} = \dfrac{\lambda^n}{k!\,\mu^n k^{n-k}}$ $\Rightarrow$ series converges if $k\mu > \lambda$

$$= \frac{k^k}{k!}\left(\frac{\lambda}{k\mu}\right)^n$$

M/M/k queue is transient if $k\mu < \lambda$

positive recurrent if $k\mu > \lambda$

null recurrent if $k\mu = \lambda$

M/M/∞ queue $\lambda_n = \lambda$

$\mu_n = n\mu$

Examine each series: $\displaystyle\sum_{n=1}^{\infty} \frac{\mu_1 \cdots \mu_n}{\lambda_1 \cdots \lambda_n} = \sum_{n=1}^{\infty} \frac{\mu(2\mu)\cdots(n\mu)}{\lambda^n} = \sum_{n=1}^{\infty} n!\left(\frac{\mu}{\lambda}\right)^n$

Ratio test: $\displaystyle\lim_{n\to\infty} \frac{(n+1)!\left(\frac{\mu}{\lambda}\right)^{n+1}}{n!\left(\mu/\lambda\right)^n} = \lim_{n\to\infty} (n+1)\frac{\mu}{\lambda} = \infty$

So series always diverges (never transient)

$$\sum_{n=1}^{\infty} \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} = \sum_{n=1}^{\infty} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n \quad \text{always converges}$$

$$= e^{\lambda/\mu} - 1$$

M/M/∞ queue is always positive recurrent, for all $\mu, \lambda > 0$

In fact, $\pi_n = e^{-\lambda/\mu} \dfrac{(\lambda/\mu)^n}{n!}$ Poisson distribution with parameter λ/μ

Mean queue length is λ/μ.